Characterization of dCDK12, hCDK12, and hCDK13 in the Context of RNA Polymerase II CTD Phosphorylation and Transcription-Associated Events

by

Bartlomiej Bartkowiak

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Arno Greenleaf, Supervisor

_____
Mariano Garcia-Blanco

_____
Kenneth Kreuzer

_____
Paul Modrich

_____
Pei Zhou

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2014

ABSTRACT

Characterization of dCDK12, hCDK12, and hCDK13 in the Context of RNA Polymerase II CTD Phosphorylation and Transcription-Associated Events

by

Bartlomiej Bartkowiak

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Arno Greenleaf, Supervisor

_____
Mariano Garcia-Blanco

_____
Kenneth Kreuzer

_____
Paul Modrich

_____
Pei Zhou

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2014

Abstract

Eukaryotic RNA polymerase II (RNAPII) not only synthesizes mRNA, but also coordinates transcription-related processes through the post-translational modification of its unique C-terminal repeat domain (CTD). The CTD is an RNAPII specific extension of the enzyme's largest subunit and consists of multiple repeating heptads with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. In *Saccharomyces cerevisiae* (*Sc*), RNAPII committed to productive elongation is phosphorylated at the S_2 positions of the CTD, primarily by CTDK-I (composed of the CDK-like Ctk1, the cyclin-like Ctk2, and Ctk3) the principal elongation-phase CTD kinase in *Sc*. Although responsible for the bulk of S_2 phosphorylation *in vivo*, Ctk1 coexists with the essential kinase Bur1 which also contributes to S_2 phosphorylation during elongation. In higher eukaryotes there appears to be only one CTD S_2 kinase: P-TEFb, which had been suggested to reconstitute the activity of both of the *Sc* S_2 CTD kinases. Based on comparative genomics, we hypothesized that the previously-unstudied *Drosophila* CDK12 (dCDK12) and little-studied human CDK12 and CDK13 (hCDK12 and hCDK13) proteins are CTD elongation-phase kinases, the metazoan orthologs of yeast Ctk1. Using fluorescence microscopy we show that the distribution of dCDK12 on formaldehyde-fixed polytene chromosomes is virtually identical to that of hyperphosphorylated RNAPII, but is distinct from that of P-TEFb. Chromatin immunoprecipitation experiments confirm that

dCDK12 is present on the transcribed regions of active *Drosophila* genes in a pattern reminiscent of a S_2 CTD kinase. Appropriately, we show that dCDK12, hCDK12, and hCDK13 purified from nuclear extracts manifest CTD kinase activity *in vitro* and associate with CyclinK, implicating it as the cyclin subunit of the kinase. Most importantly we demonstrate that RNAi knockdown of dCDK12 in *Drosophila* cell culture and hCDK12 in human cell lines alters the phosphorylation state of the CTD. In an effort to further characterize the transcriptional roles of human CDK12/CyclinK we overexpress, purify to near homogeneity, and characterize, full-length hCDK12/CyclinK. Additionally, we also identify hCDK12 associated proteins via mass spectrometry, revealing interactions with multiple RNA processing factors, and attempt to engineer an analog sensitive CDK12 human cell line. Overall, these results demonstrate that CDK12 is a major elongation-associated CTD kinase, the ortholog of yCtk1. Our findings clarify the relationships between two yeast CDKs, Ctk1 and Bur1, and their metazoan homologues and draw attention to major metazoan CTD kinase activities that have gone unrecognized and unstudied until now. Furthermore, the results suggest that hCDK12 affects RNA processing events in two distinct ways: Indirectly through generating factor-binding phospho-epitopes on the CTD of elongating RNAPII and directly through binding to specific factors.

Dedication

I would like to dedicate this dissertation to Ella, my wonderful wife; my father and mother, who spurred my interest in science and whose support was invaluable not only through graduate school but throughout my entire life; my mentor, Dr. Arno Greenleaf; and my family and friends.

Contents

List of Tables

List of Figures

List of Abbreviations

AA	Amino Acid
AP	Affinity Purified
ChIP	Chromatin Immunoprecipitation
CTD	C-Terminal Domain of RNA polymerase II
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Dm	*Drosophila Melanogaster*
FVP	Flavipiridol
GST	Glutathione S-Transferase
GST-yCTD	GST and full length yeast CTD (26 repeats) fusion protein
MW	Molecular Weight
NHEJ	Non-Homologous End Joining
ORF	Open Reading Frame
PAM	Protospacer Adjacent Motif
PIC	Pre-Initiation Complex
Poly(A)	Poly-Adenylation
RNAPII	RNA Polymerase II
Sc	*Saccharomyces cerevisiae*
Ser2/Ser5/Ser7	Serine 2, 5, or 7 positions within the CTD heptad

Ser2P/Ser5P/Ser7P	Phosphorylated Serine 2, 5, or 7 positions with the CTD heptad
SH2	Src Homology 2
Sp	S*chizzosaccharomyces pombe*
SRI	Set2 Rbp1 Interacting
TSS	Transcription Start Site
TTS	Transcription Termination Site

Acknowledgements

There are many people that contributed directly and indirectly to the completion of this dissertation. Most importantly I would like to thank Dr. Arno Greenleaf for his expertise, guidance, and for being an understanding and patient mentor; it may be obvious but it bears repeating that without him none of this would have been possible. In the same vein I would like to thank the members of the Greenleaf lab, both past and present, especially Dr. Pengda Liu, Dr. Tiffany Sabin-Winsor, Dr. April MacKellar, Chris Yan, Dr. Jan Jones, and Dr. Hemali Phatnani. April should be especially applauded as she dealt with me as a rotation student (I am sure that I drove her up the wall). I would also like to thank Dr. Craig Bennett for his suggestions, sense of humor, and company; he will be missed. I am also indebted to my thesis committee for their time and effort in supporting me throughout this process and for their suggestions and encouragement. I would especially like to thank Dr. Mariano Garcia-Blanco for including us in the RNA Center and for his support and mentorship. The RNA Center open lab space has been a wonderful environment in which to do research and I would like to acknowledge the RNAi facility and the members of the Garcia-Blanco lab (again both past and present) for their willingness to contribute both scientific and non-scientific advice and for their impressive ability to have a good time; a special thanks to Dr. Jason Somarelli and Dr. Daneen Schaeffer, you really helped make this a fantastic experience. I must also give another special thank you to Dr. Shelton Bradrick for his experimental advice and to

Nicholas Barrows for introducing me to the CRISPR/Cas system. Another big thank you is owed to Dr. Rasheed Gbadegesin for his support and for his role in my journey to graduate school. Finally I must thank Dr. Karen Adelman and Dr. Dan Gilchrist at NIEHS for being kind enough to teach me the ins and outs of ChIP along with various genome wide techniques and for sharing copious amounts of their reagents, expertise, and time.

Last but not least I would like to thank everyone who has helped me on a more personal level. My family has been a constant source of support; especially my parents, Magda and Mirek, and my wife, Ella; they have bravely endured my various idiosyncrasies and kept me grounded through this process. Finally, thank you to all of the friends I have made during this time and for the wonderful times we had together. A truly special thanks to Dr. Nam Tonthat whose early morning carpool was key to much of my productivity and proved to me once and for all that I am not a morning person.

1. Introduction[1]

1.1 The C-terminal Domain of RNA Polymerase II

Eukaryotic transcription and the concomitant pre-mRNA processing that accompanies it require the precise coordination between, and recruitment of, specific sets of factors at specific stages of the transcription cycle. This coupling of transcription and associated processes has been shown to be dependent on a particular feature of RNA Polymerase II (RNAPII), the C-terminal repeat domain or CTD (Bartkowiak et al., 2011; Corden, 2013; Phatnani and Greenleaf, 2006). Distinguishing RNAPII from its prokaryotic and eukaryotic (RNAPIII and RNAPI) counterparts, the CTD is an unstructured extension of the enzyme's largest subunit, Rpb1, and is composed of a tandem array of seven amino acid repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. The number of these heptad repeats varies from organism to organism and appears to correlate with genomic complexity; there are 26 repeats in yeast, 44-46 in *Drosophila*, and 52 in humans (Allison et al., 1988; Corden, 1990). Despite being dispensable for the catalytic activity of RNAPII, the CTD is conserved throughout evolution and is essential to life; no animal, plant, or fungal Rpb1 sequence found to date lacks a CTD (Stiller and Hall, 2002). Additionally, the total number of heptad repeats is also important; as an example, yeast cells or fruit files that have two thirds of the CTD repeats removed are

[1] Adapted from Bartkowiak, B., Mackellar, A. L., Greenleaf, A. L., 2011. Updating the CTD Story: From Tail to Epic. Genet Res Int. 2011, 623718.

not viable (Nonet et al., 1987; Zehring et al., 1988). Somewhat surprisingly, despite the overall conservation of the CTD the range of deviation from the consensus sequence can be highly variable (Chapman et al., 2008). The amount of variation depends on the organism in question; for a large group of eukaryotes called the "CTD-clade," the consensus sequence is strongly stabilized (Stiller and Hall, 2002). But even within this clade some organisms stand out as exceptions, an example of particular interest being *Drosophila melanogaster* which contains only two canonical CTD consensus repeats. Despite the nonconformance of *Drosophila* at the heptad level, the overall repetitive structure of the CTD is still conserved as 75% of the sequence still lies in the tandemly repeated register (Stiller and Hall, 2002) (Fig. 1.1).

The conservation and expansion of the CTD throughout evolution suggest that it plays a fundamental, but unknown role in RNAPII transcription. Initial models for CTD function were varied and included the CTD interacting with initiation factors and/or DNA, removing DNA bound proteins, restricting RNA secondary structure, and even localizing RNAPII to specific locations in the nucleus (Corden, 1990). Many of these models, especially the ones postulating protein-protein interactions, recognized the potential of CTD post-translational modification, especially phosphorylation, as a regulatory mechanism for CTD function.

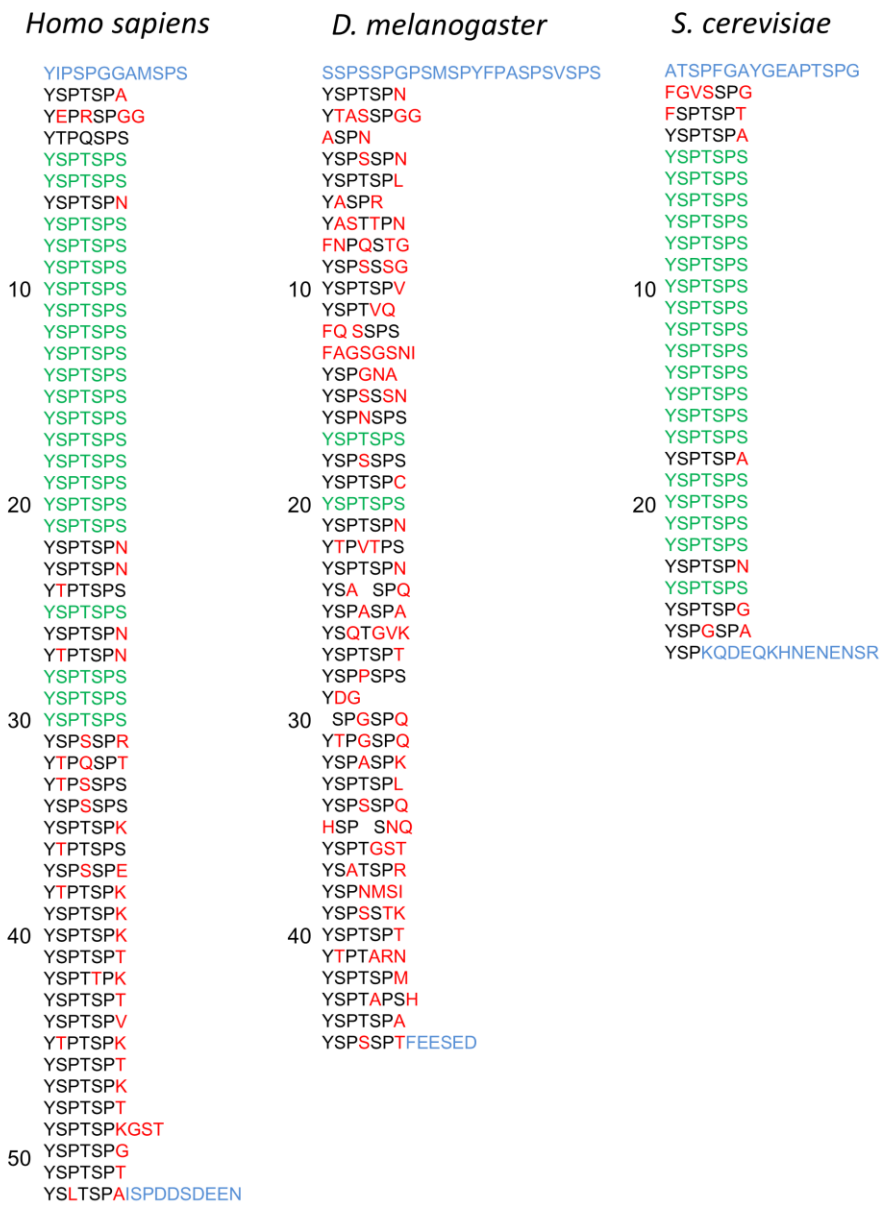

Figure 1.1: Amino acid sequence of human, *Drosophila melanogaster*, and *Saccharomyces cerevisiae* Rpb1 C-terminal domains. Residues that conform to the consensus sequence are shown in black (or if the entire heptad contains canonical sequence, highlighted in green); residues that deviate from the consensus are in red; and terminal/upstream residues that do not conform to the consensus are shown in blue. Sequences were obtained from the UCSC Genome Browser (Kent et al., 2002).

The idea that phosphorylation plays a regulatory role in transcription was first proposed in the 1970s. Early research on RNAPII revealed that the polymerase underwent phosphorylation *in vivo* (Bell et al., 1976), a modification that resulted in the manifestation of Rpb1 as several distinct ,by apparent molecular weight, species on SDS PAGE gels: A nonphosphorylated IIa form (215kD), a phosphorylated IIo form (240kD), and a proteolyzed IIb form which lacked the CTD (180kD) (Dahmus, 1994; Kedinger et al., 1974; Schwartz and Roeder, 1975). It was also noted that the IIa, but not the IIb form could be further phosphorylated by incubation with casein kinase (Dahmus, 1981); this insight, coupled with molecular cloning techniques, resulted in both the discovery of the CTD and the realization that it is the vector of RNAPII phosphorylation (Allison et al., 1985; Corden et al., 1985; Dahmus, 1994). Although exposed in terms of its sequence and potential modification, the CTD's true function would continue to be a mystery for almost another ten years.

The discovery of mRNA processing CTD binding proteins in the mid-1990s (Corden and Patturajan, 1997; Yuryev et al., 1996), coupled with the fact that newly synthesized mRNA crosslinked exclusively to the hyperphosphorylated form of the CTD (Bartholomew et al., 1986; Cadena and Dahmus, 1987; Dahmus, 1994), supported the proposal that the CTD linked transcription to transcription associated processes (Greenleaf, 1993). This proposal has been experimentally verified over the last three decades and we now know that the CTD functions as a selective binding scaffold for

nuclear factors, allowing it to coordinate transcription with varied nuclear processes in a temporal fashion (Phatnani and Greenleaf, 2006).

1.2 The "phospho-CTD cycle" and the "CTD code"

1.2.1 Phosphorylation of the CTD

The binding specificity of the CTD, and therefore the recruitment of particular factors, is mainly determined by the CTD's phosphorylation state, which undergoes a series of changes throughout the transcription cycle. The primary structure of the CTD heptad repeat lends itself as a fantastic target for this particular post translational modification as five out of seven of its residues can be phosphorylated. Serine 2, Serine 5, and Serine 7 of the heptad repeat have been identified as the key players in this transcriptionally regulated phosphorylation; however, recent studies have also demonstrated the phosphorylation and involvement of Threonine 4 and Tyrosine 1 (Hintermair et al., 2012; Hsin et al., 2014; Hsin et al., 2011; Mayer et al., 2012).

In a highly generalized model of the "phospho-CTD cycle," RNAPII is recruited for assembly at the promoter with an unphosphorylated CTD; CTD phosphorylation prior to preinitiation complex (PIC) formation has an inhibitory effect on transcription initiation (Hengartner et al., 1998). Upon PIC formation, the CTD is phosphorylated at the Ser5 and Ser7 positions; this is followed by an increase in Ser2 phosphorylation during elongation (Komarnitsky et al., 2000), yielding a CTD that contains a mix of doubly (and when considering Ser7, perhaps triply) phosphorylated repeats at Ser2 and

Ser5 in the center of the gene (reviewed in (Bartkowiak et al., 2011; Eick and Geyer, 2013; Phatnani and Greenleaf, 2006) and (Buratowski, 2009)). As the polymerase elongates towards the 3' end of the gene, the activity of Ser5-specific phosphatases decrease the Ser5 phosphorylation levels, leaving CTDs phosphorylated at Ser2 to terminate transcription. However the CTD of terminating RNAPII may also be phosphorylated at Ser7 positions, as the Ser7 mark has been reported to be retained at high levels throughout the entire transcription unit on many protein-coding genes (Tietjen et al., 2010). In the first of many complications to this simplified model, we must note the phosphorylation of Tyr1 (Baskaran et al., 1993) and Thr4 (Zhang and Corden, 1991), which having been recently characterized as functionally significant (Hintermair et al., 2012; Hsin et al., 2014; Hsin et al., 2011; Mayer et al., 2012), are now being incorporated into the general model (Fig. 1.2).

Despite being highly intuitive, and very useful, this segmented (Ser5P to Ser2/5P to Ser2P with Ser7P throughout) description of the "phospho-CTD cycle" is oversimplified for several reasons. First, although this model accounts for variations in

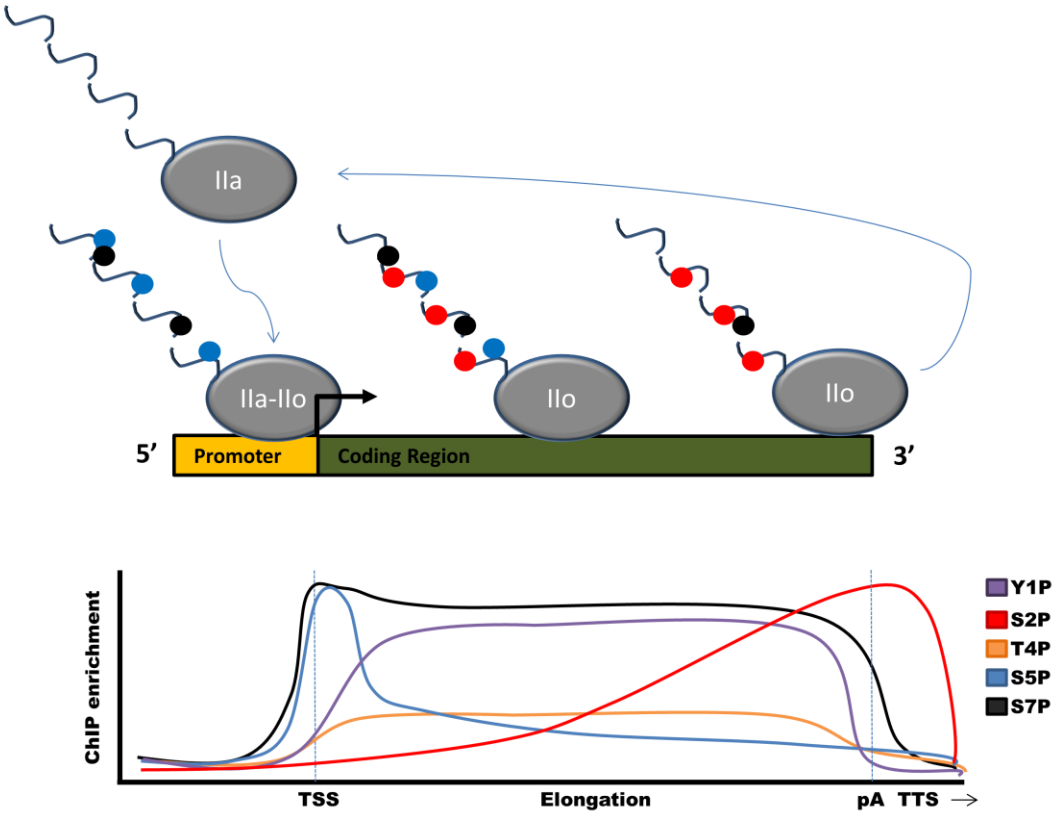

Figure 1.2: Simplified schematics of the "phospho-CTD cycle" and CTD modifications as determined by Chromatin Immunoprecipiatation (ChIP) experiments in *Saccharomyces cerevisiae*. Unphosphorylated RNAPII (IIa) is recruited to promoters, where upon PIC formation the CTD (grey squiggles) is phosphorylated at the Ser5 (blue) and Ser7 (black) positions. During elongation there is an increase in Ser2 (red) phosphorylation with a concomitant decrease of Ser5 phosphorylation; at the 3' end of the transcription unit Ser2 phosphorylation predominates. The ChIP visualization of the CTD modifications also displays the enrichment pattern of the more recently characterized phosphorylation of Tyr1 and Thr4. ChIP signals are not normalized to RNAPII levels. The ChIP enrichment view is adapted from (Eick and Geyer, 2013).

the general phosphorylation pattern of the CTD, it fails to fully capture the highly dynamic nature of the process. The phosphorylation state of the CTD is likely to be in continuous flux throughout the entire transcription cycle, with multiple kinases and phosphatases working together to maintain specific phosphorylation patterns on a particular subsets of heptads. For example, it has been shown that the Ser7 phosphates

at the 3' end of the gene are actually placed anew, after being removed by a phosphatase (Tietjen et al., 2010). Another important set of complications to the simple gradient model is the existence of specific patterns of phosphorylated repeats within the CTD. It is worth noting that the repetitive nature of the CTD makes it impossible to determine exactly which of the 52 heptad positions (in mammals) are phosphorylated. In addition the particular pattern of phosphorylated residues within each heptad and within sequential heptad units is also unknown; therefore the Ser2, 5, and 7 phosphorylated CTD of elongating RNAPII is likely to be composed of heptads phosphorylated at one, both, all, and none of the relevant positions. The general model of the "phospho-CTD cycle" also fails to account for exceptions to the canonical patterns of phosphorylation, the differences among the modifications of the CTD as RNAPII transcribes specific genes, and is further complicated by the existence of other CTD modifications.

1.2.2 Other CTD modifications

It has not escaped notice that the modification of the CTD is not theoretically limited to phosphorylation. Many other post translational modifications have been observed, including glycosylation (Kelly et al., 1993; Ranuncolo et al., 2012) , acetylation (Schroder et al., 2013), and methylation (Sims et al., 2011). Non-canonical modifications, once fully characterized as functionally significant, have the potential to expand the CTD code further and redefine aspects of the general model.

One important CTD modification, which until recently had not been directly observed to play a direct role in factor binding, is the enzymatic isomerization of the peptidyl-proline bonds in the heptad repeats. Although multiple studies have suggested a role for the CTD interacting peptidyl-prolyl cis/trans isomerases (Ess1 in yeast and Pin1 in humans) in transcription and CTD phosphorylation (Krishnamurthy et al., 2009; Wilcox et al., 2004; Xu and Manley, 2007), all of the structures of CTD-substrates/CTD-binding protein complexes revealed the CTD proline residues to be exclusively in the more energetically stable, and therefore predominant, trans state. This changed when two structural studies found that the Ser5 specific CTD phosphatase Ssu72 bound to the cis conformation of a Ser5-Pro6 motif within the heptad repeat (Werner-Allen et al., 2011; Xiang et al., 2010). Concordantly, the activity of the proline isomerase Ess1 was found to facilitate the rapid dephosphorylation of the CTD by Ssu72 *in vitro*, suggesting that this cis/trans interconversion plays a role in the fine-tuning of the phosphorylation state of the CTD (Werner-Allen et al., 2011). These findings have broad implications for CTD biology, both in increasing the number of distinct CTD states and serving as a regulatory mechanism for CTD phosphorylation. However, it still remains to be determined whether proline isomerization is a general property of RNAPII transcription or is gene specific (Werner-Allen et al., 2011), this distinction may apply to other types of modifications as well.

As previously mentioned, the general "phospho-CTD cycle" fails to account for gene specific patterns and modifications. Two good examples of non-canonical CTD modifications which affect a specific transcript class are the methylation of an arginine (R1810) in heptad 31 (Sims et al., 2011) and the acetylation of the eight lysines in the human CTD (see Fig. 1.1). Mediated by the methyltransferase CARM1 and inhibited by Ser5 and Ser2 phosphorylation, methylation of the CTD appears to repress the expression of snRNAs and snoRNAs in a general manner (Sims et al., 2011). The acetylation of the lysines is mediated by p300/KAT3B, and despite the fact that it does not affect the expression of housekeeping genes, mutation of the lysines or inhibition of p300 results in the loss of epidermal growth factor induced expression of some immediate-early genes such as c-Fos and Egr2 (Schroder et al., 2013). These and other modifications of the non-canonical heptads may serve as a discriminatory mark for RNAPII recruited to particular genes or transcript classes. It should also be noted that Ser7P is thought to be transcript class specific CTD modification as Ser7 to alanine mutations in the CTD cause a defect in snRNA transcription while having little effect on protein coding genes (Egloff et al., 2007). However, the ubiquitous nature of Ser7P on protein coding genes, along with the finding that Ser7 is enriched on RNAPII within introns (Kim et al., 2010) argues for some, perhaps more subtle, functional role for Ser7P on most transcription units.

Thus the general phospho-CTD cycle has given way to a "CTD code" of staggering complexity; one that we are just beginning to explore in detail. This complexity reflects the vast number of different genes, processing events, and transcriptional programs that RNAPII must coordinate.

1.3 Functions of the CTD

The CTD has been implicated in a broad spectrum of transcription associated functions and its collection of binding partners has continued to expand over the last few years. Important associated processes include mRNA (and snRNA) capping, splicing, 3'end processing, termination, and more recently nuclear export. In terms of non-RNA processing associated events, the CTD has been shown to play roles in transcriptional activation, cotranscriptional chromatin modification, chromatin remodeling, and genome stability[2]. Of course the analysis of the CTD's role in each specific function is very challenging, as many of the CTD-mediated transcriptional processes are interlinked. For example, capping has been shown to influence both the splicing of the first intron and 3' end processing (Cooke and Alwine, 1996; Flaherty et al., 1997; Hart et al., 1985; Inoue et al., 1989; Lewis et al., 1996); splicing of the last intron affects 3' end processing, and vice versa (Bauren et al., 1998; Niwa and Berget, 1991; Vagner et al., 2000); and alternative splice site choices are affected by the co-

[2] See below and (Bartkowiak et al., 2011; Corden, 2013; MacKellar and Greenleaf, 2011; Winsor et al., 2013) for details.

transcriptional histone modifications at splice site junctions (Allo et al., 2009; Luco et al., 2010; Schor et al., 2009). While these interactions demonstrate the high degree of coordination involved in mRNA synthesis, they unfortunately complicate the interpretation of functional studies. In a broad sense, it is fair to state that exactly how Ser2, Ser5, and Ser7 phosphorylation affect initiation, elongation, and termination remains poorly understood. Despite the lack of a universal synthesis, particular aspects of CTD function have been characterized to an impressive level of detail and many of the more enigmatic functions are slowly giving way to continued investigation. I will go over a few of the CTD's most prominent functions in the following subsections, taking the opportunity to also survey some of the key factors which bind to the phospho-CTD.

1.3.1 5' Capping

One of the most clearly recognized functions of the CTD is its involvement in the 5' end capping of mRNA through the recruitment of the capping machinery. The modification of the 5' end of the RNA with the 5' 7-methyl guanosine cap is unique to RNAPII transcripts, and occurs just after the transcript clears the polymerase's exit channel (Jove and Manley, 1984; Rasmussen and Lis, 1993). Transcripts made by a CTD-less RNAPII were found to be inefficiently capped, leading to the characterization of the physical interaction between the capping enzymes and the phospho-CTD (Cho et al., 1997; McCracken et al., 1997a; Yue et al., 1997). Subsequent studies showed that the capping enzyme associates with the 5' end of genes *in vivo*, and that this association is

dependent on the phosphorylation of the CTD (Komarnitsky et al., 2000; Schroeder et al., 2000). This localization correlates nicely with the enzyme's function and is dependent on the phosphorylation of Ser5 (Kanin et al., 2007; Schroeder et al., 2000). In addition to raising the local concentration of the capping enzyme near the exit channel, the CTD has also been shown to stimulate its activity: An example of a phosphorylation state specific function, mammalian guanylyltransferase was found to bind to both Ser2P and Ser5P synthetic heptad repeats, but was allosterically activated only by Ser5P (Ho and Shuman, 1999). The interaction between the 5' capping machinery and the CTD has also been investigated structurally and these studies resulted in some interesting insights. The crystal structure of the *C. albicans* guanylyltransferase Cgt1 and a synthetic Ser5P four heptad repeat revealed that the CTD peptide binds to an extended docking site on the enzyme surface using two non-adjacent heptads, which leaves a full heptad repeat looped out away from the interaction site (Fabrega et al., 2003). This looping not only demonstrates the inherent flexibility of the CTD, it also suggests that by binding two remote heptads, CTD binding factors may be able to loop out large portions of the CTD. The loop could potentially result in the formation of novel structural motifs, which could in turn serve as binding sites for other CTD binding factors, leading to an organized, sequential binding (Meinhart et al., 2005). Whether this actually occurs is still on open question; however, recent studies have shown that the binding of some well-

known 3′ end processing factors to the CTD appears to be cooperative in nature (Lunde et al., 2010).

1.3.2 3' End processing

Another well recognized function of the CTD is its role in 3′ end processing and termination[3]. Analogous to capping, transcription by a CTD less RNAPII was shown to affect both processes, and cleavage and polyadenylation factors were found to bind to the phospho-CTD (Barilla et al., 2001; Hirose and Manley, 1998; Kyburz et al., 2003; Licatalosi et al., 2002; Maniatis and Reed, 2002; McCracken et al., 1997b; Proudfoot et al., 2002). Accordingly, the deletion of Ctk1 (a Ser2 CTD kinase) in yeast has been shown to decrease the efficiency of cleavage at poly(A) sites (Skaar and Greenleaf, 2002) and result in the disruption of polyadenylation factor recruitment to the 3′ end of the gene (Ahn et al., 2004). Genome wide analysis has also shown that depletion of Ctk1 using a tetracycline-repressible degron mutant, causes a "pileup" of polymerases at the poly(A) site in a subset of genes with good consensus poly(A) sequences (Kim et al., 2010). This increase in RNAPII occupancy suggests that improper CTD phosphorylation at these sites can result in a strong transcriptional pause; perhaps due to the rate limiting recruitment of a specific factor. Strongly linked to polyadenylation, termination has also been reported to be affected by the CTD (McCracken et al., 1997b); in addition Rtt103, a component of the termination complex has been shown to bind Ser2P CTD (Kim et al.,

[3] See (Proudfoot, 2004) and (Buratowski, 2005) for more information.

2004). However the role of CTD modification in termination is not yet well understood[4]. Intriguingly, it has been observed that the recruitment of one well-recognized CTD binding 3' end processing factor, Pcf11 (which preferentially binds to Ser2 phophorylated CTD) does not directly correlate with the level of Ser2 CTD phosphorylation (Ahn et al., 2004; Licatalosi et al., 2002; Mayer et al., 2010); analysis by chromatin immunoprecipitation (ChIP) indicates that Pcf11 is recruited mainly at the poly(A) site, while Ser2P levels rise throughout the coding region. Potential explanations for this phenomenon highlight some of the interesting complications surrounding phospho-CTD factor recruitment. Perhaps Pcf11 requires a certain threshold of Ser2P or both Ser2P and an external signal, such as the presence of the newly synthesized polyadenylation site, the unmasking of particular CTD epitopes, or the presence of additional factors (Pcf11 CTD binding was recently shown to be cooperative (Lunde et al., 2010)). Other considerations include the remodeling of the CTD via a pattern specific change, such as the formation of Ser2P only heptads, or modifications undetectable to ChIP. With regards to the latter and the previous discussion on proline isomerization, Pcf11 is reported to specifically recognize three neighboring trans prolines within a mixed population of cis-trans isomers (Noble et al., 2005). It is likely that the recruitment of many other CTD binding factors is also

[4] See (Buratowski, 2005), (Buratowski, 2009), and (Kuehner et al., 2011) for further discussion.

mediated through multiple mechanisms and dependent on the satisfaction of particular sets of conditions.

1.3.3 snRNA Processing

One of the more recently discovered functions of the CTD is its role in snRNA transcription and 3′ end processing (Jacobs et al., 2004; Medlin et al., 2003). Even though they are transcribed by RNAPII, snRNAs are unlike most coding genes; they do not undergo splicing or polyadenylation, instead relying on a conserved 3′ box RNA processing element downstream of the coding region for proper 3′ end processing and termination (Egloff et al., 2008). The 3′ end processing of snRNAs has received a lot of attention as it is currently the only specific function attributed to the phosphorylation of the Ser7 position of the heptad repeat (Egloff et al., 2007). Seemingly dispensable for viability and the expression of protein coding genes, Ser7 has been found to be essential for endogenous snRNA gene expression. This requirement for Ser7 phosphorylation was subsequently linked to the Integrator complex (Egloff et al., 2007); a large CTD associated multiprotein complex involved in snRNA 3′ end processing (Baillat et al., 2005). Further characterization of the CTD-Integrator interaction not only revealed that both Ser2P and Ser7P were required, but that efficient binding required a specific arrangement of the modifications (Egloff et al., 2010). Screening of synthetic diheptad repeats revealed that although maximal binding was achieved with both Ser2 and Ser7 repeats phosphorylated, the minimal interaction domain consisted of a Ser7P on the first

heptad followed by a Ser2P on the second heptad; any other combination of two phosphates was insufficient for Integrator binding (Egloff et al., 2010). These findings lend further support for the pattern specific binding of CTD associated factors and, coupled with the fact that Integrator is specific to snRNA genes, also suggest that the appropriate Ser2P/Ser7P patterns may be snRNA gene specific (although many other discriminatory mechanisms could be at play, an example might be the previously discussed methylation of R1810).

1.3.4 Histone modification

Tying together two important aspects of gene expression and transcriptional coordination, the CTD has also been shown to be involved in the cotranscriptional modification and remodeling of histones. Providing the first clear link between the CTD and histone modification, the Set1 and Set2 histone methyltransferases were found to be recruited to actively transcribed genes at specific stages of the transcription cycle (5′ end vs interior of the gene respectively) through an interaction with the Ser5P and Ser2/5P CTD (Kizer et al., 2005; Krogan et al., 2003a; Krogan et al., 2003b). Thus Set1 activity, the methylation of histone H3 at the K4 position, peaks near the promoter, while Set2 activity, methylation of H3 at K36, occurs downstream in the coding region. The H3K36 trimethylation mark can be used to identify transcriptionally active genes and has been shown to suppress inappropriate transcription from cryptic promoters through what appeared to be the recruitment of the histone deacetylase Rpd3S (Joshi and Struhl, 2005;

Keogh et al., 2005). Surprisingly, more recent studies have found that Rpd3S is actually recruited via direct binding to the phospho-CTD; although H3K36 trimethylation is dispensable for Rpd3S recruitment, it appears to be required for activation of its deacetylation activity (Drouin et al., 2010; Govind et al., 2010). Intriguingly Set2 is capable of H3K36 dimethylation independent of its CTD interacting SRI domain or Ctk1 (Youdell et al., 2008), implying that only one specific aspect (H3K36 trimethylation) of Set2's activity is regulated by interaction with the phospho-CTD. Another interesting chromatin related CTD binding factor is the histone H3 chaperone and transcription elongation factor Spt6 (Yoh et al., 2007). Spt6 is a nucleosome remodeler and has been shown to bind the Ser2P CTD through a tandem SH2 domain (Burugula et al., 2014; Sun et al., 2010). Spt6 has been found to be required for the optimal loading of the mammalian Set2 (HYPB/SETD2) on the coding regions of several genes (Yoh et al., 2008), thus linking nucleosome reassembly with elongation coupled H3K36 trimethylation *in vivo*. Several other chromatin related CTD interactions include the recruitment of the chromatin remodeling factor CHD8 (Rodriguez-Paredes et al., 2009) and the FACT histone chaperone via HP1 (Kwon et al., 2010).

1.3.5 Splicing

One of the more intriguing, but relatively poorly characterized, functions of the CTD is its involvement in cotranscriptional splicing (Goldstrohm et al., 2001; Munoz et al., 2010). Although neither active transcription nor the CTD are absolutely required for

splicing (presynthesized mRNAs can be spliced by injection into Xenopus oocytes or by incubation with nuclear extracts (Bird et al., 2004)), experimental evidence accumulated over the last three decades implicates the CTD as a key player in the coupling between the two processes. The link between the CTD and splicing was first noted in the mid to late 90s with the observation that the hyperphophorylated RNAPII associates with the SR (Serine/Arginine rich) family of splicing factors and with components of the splicing machinery (Kim et al., 1997; Mortillaro et al., 1996; Yuryev et al., 1996). Concordantly transcription by a CTD less RNAPII was shown to result in low splicing efficiency *in vivo* (McCracken et al., 1997b) and the addition of anti-CTD antibody or exogenous expression of CTD peptides resulted in the accumulation of unspliced transcripts (Yuryev et al., 1996) and the nuclear reorganization of splicing factors (Du and Warren, 1997). In what could be considered the reciprocal experiment, it was also shown that isolated CTD fragments and purified phosphorylated RNAPII were able to activate splicing reactions *in vitro* (Hirose et al., 1999; Zeng and Berget, 2000).

In addition to the SR-like CTD associated factors (SCAFs) (Patturajan et al., 1998b; Yuryev et al., 1996), several other CTD binding splicing factors have been identified, including the yeast U1 snRNP component Prp40 (Morris and Greenleaf, 2000) and the mammalian splicing factors CA150 (TCERG1)(Carty et al., 2000), PSF, and p54/NRB (Emili et al., 2002). A recent addition has been the splicing factor U2AF, which was shown to be recruited to the Ser2P CTD in complex with another splicesome

component, Prp19 (David et al., 2011; Gu et al., 2013). Although alternative splicing has been directly demonstrated to be affected by the presence of the CTD (de la Mata and Kornblihtt, 2006), due to the fact that RNAPII elongation rate has also been shown to play a role in alternative splicing (de la Mata et al., 2003; Kadener et al., 2001) it is difficult to determine how much of the effect is due to the specific recruitment of splicing factors to the CTD and how much is due to changes in the elongation rate (kinetic coupling) or other processes. With regards to kinetic coupling and the CTD, one study which merits specific mention is that of Batsche et al. who reported that the inclusion of a set of alternative exons in the middle of the CD44 gene was dependent on the accumulation of RNAPII induced by the catalytic subunit of the SWI/SNF chromatin remodeling complex (Batsche et al., 2006). Surprisingly, this "stalled" RNAPII exhibited a SWI/SNF-dependent switch of the RNAPII phosphorylation state from the elongation characteristic Ser2P to the more promoter characteristic Ser5P, perhaps creating a barrier to further elongation. In addition to its general role in the recruitment of splicing factors, it has also been suggested that the CTD may function as a molecular tether for distant splice sites within the mRNA. This would be accomplished through the binding of the nascent 3′ splice site to a CTD associated splicing factor, thus effectively immobilizing it near the polymerases mRNA exit channel in anticipation of the cognate 5′ splice site. Enhancement of the local concentration of the splice sites would then dramatically increase the efficacy of the splicing reaction (Goldstrohm et al., 2001). This

tethering model is especially attractive in higher eukaryotes in which the intron lengths of many genes exceed thousands of base pairs. Although studies have demonstrated that mRNA tethering is likely (Dye et al., 2006), it has yet to be directly linked to the CTD and its phosphorylation.

1.4 The CTD kinases

The proper regulation of CTD associated processes is dependent on the appropriate modification of the CTD during the transcription cycle. The specific phosphorylation events within the heptad repeat are mediated through the activity of a transcription-associated subset of cyclin-dependent kinases (CDKs) known as CTD kinases. Unlike their cell cycle counterparts, the CTD kinases form complexes with members of the non-cycling "transcription cyclin" family and are active throughout the cell cycle. Nevertheless, CTD kinase activity is tightly regulated by a variety of mechanisms, including selective recruitment, binding by kinase-associated factors, and sequestration by inhibitory factors. Although somewhat promiscuous *in vitro* (for example the "Ser2-specific" CTD kinase Ctk1 can phosphorylate both Ser2 and Ser5 *in vitro* (Jones et al., 2004), *in vivo* the CTD kinases are selective for particular residues within the heptad repeat and stages of transcription. Thus the various CTD kinases are most conveniently presented in the context of a segmented transcription cycle; however, it should be made clear that in reality the various kinase activities lack the clearly delineated boundaries that such a presentation suggests. It should also be kept in mind

that the CTD kinases function in conjunction with the slightly more enigmatic CTD phosphatases, several of which have been characterized in yeast: Fcp1 which targets Ser2P (Cho et al., 2001), Ssu72 which targets S5P and has been implicated in targeting S7P (Krishnamurthy et al., 2004; Reyes-Reyes and Hampsey, 2007; Zhang et al., 2012) , and the controversial Rtr1 which targets S5P (deletion of Rtr1 reduces S5P levels *in vivo*, but *in vitro* experiments have failed to identify an Rtr1 CTD phosphatase activity implying that Rtr1 has a noncatalytic/structural role in CTD dephosphorylation) (Mosley et al., 2009; Ni et al., 2014; Xiang et al., 2012).

1.4.1 Initiation and the promoter – the serine 5 position and CDK7

The phosphorylation of CTD Ser5 and Ser7 residues during the formation of the preinitiation complex is mediated by the CTD kinase subunit of the general transcription factor TFIIH: Kin28/Ccl1 in yeast and CDK7/CyclinH in metazoa (Akhtar et al., 2009; Glover-Cutter et al., 2009; Kim et al., 2009; Komarnitsky et al., 2000; Schroeder et al., 2000); see (Fig. 1.4) for clarification on the CTD kinases in various organisms. In an elegant interplay, the kinase activity of Kin28 is stimulated by the mediator coactivator complex which binds to, and delivers, unphosphorylated RNAPII to the promoter (Guidi et al., 2004). The resulting phosphorylation of the CTD leads to the dissociation of mediator (Max et al., 2007); thus after fulfilling its function, mediator is able to use the CTD and Kin28 to induce its own release from transcriptionally active RNAPII. Once thought to be essential for promoter clearance, the activity of Kin28 has been shown to

be dispensable for global gene transcription (Hong et al., 2009; Kanin et al., 2007). Despite this, Kin28 has been found to enhance polymerase progression through long genes in yeast (over 2kb), suggesting that it plays a role in transcriptional elongation or in the inhibition of premature termination (Kim et al., 2010). CDK7 also takes part in the phosphorylation and activation of other CDKs (Fisher, 2005); however, its (and Ser5P's) most clearly defined transcriptional role is the recruitment of the 5' end capping machinery. Not only does this ensure the proper processing of the nascent mRNA, it has also been shown to mediate the recruitment of the Ser2 CTD kinases in some organisms (either directly or through recruitment of the capping machinery) (Pei et al., 2006; Qiu et al., 2009; Viladevall et al., 2009); this suggests that phosphorylation of Ser5 plays a role in triggering the onset of Ser2 phosphorylation.

1.4.2 Elongation – the serine 2 position CTD kinases

Subsequent to Kin28 activity at the promoter, phosphorylation of Ser2 of the CTD heptad occurs downstream of the transcription start site (TSS) and coincides with RNAPII entry into productive elongation. Coupled with the activity of Ser5P-specific phosphatases this leads to a transition from high Ser5P to high Ser2P (as characterized by ChIP). The Ser5P to Ser2P crossover point, defined as the point at which the ChIP signals for the two CTD marks cross, is on average ~450 bp downstream of the TSS and appears to be independent of the overall gene length (Kim et al., 2010). This implies that the dynamics of Ser2 and Ser5 phosphorylation are not scaled to gene length; however

the significance of the crossover point in terms of the actual phosphorylation state of the CTD is obscure[5].

In *Saccharomyces cerevisiae,* the phosphorylation of Ser2 is primarily mediated by CTDK-I, a three subunit enzyme (consisting of Ctk1, a CDK homologue; Ctk2, a cyclin homologue; and Ctk3, whose function is unknown) (Lee and Greenleaf, 1991; Sterner et al., 1995). Although it is responsible for the bulk of Ser2 phosphorylation *in vivo,* Ctk1 is not essential for viability under standard growth conditions, or for transcriptional elongation. The CTD kinase activity of Ctk1 has been linked to several transcription-associated processes, including the recruitment of the Set2 histone methyltransferease (Xiao et al., 2003; Youdell et al., 2008), 3′ end processing (Ahn et al., 2004; Kim et al., 2010; Skaar and Greenleaf, 2002) , splicing (Phatnani et al., 2004), termination (reviewed in (Buratowski, 2009)), and the termination of small non-coding RNAs (Lenstra et al., 2013). In addition to its role as a CTD kinase, Ctk1 (independent of its kinase activity) has also been shown to be involved in the dissociation of basal transcription factors from RNAPII (Ahn et al., 2009). Despite having a principal role in CTD Ser2 phosphorylation, Ctk1 is not the only Ser2 kinase in yeast; it coexists with the essential Bur1 kinase (which consists of the CDK homologue Bur1 and the cyclin Bur2) (Yao et al., 2000). While it has been proposed that Bur1's primary transcription-related substrate is the elongation

[5] All of these conclusions depend on signals from particular antibodies via ChIP and there are significant ambiguities inherent in these signals, see (Phatnani et al., 2006) and (Bartkowiak et al., 2011) for further discussion).

factor Spt4/5(Keogh et al., 2003; Pei and Shuman, 2003), rather than the CTD, more recent evidence indicates that Bur1 binds to the Ser5P CTD and contributes to Ser2 phosphorylation during early elongation, possibly stimulating subsequent Ctk1 activity (Liu et al., 2009; Qiu et al., 2009). Bur1 has also been found to exhibit an elongation phase Ser7 kinase activity which appears to counteract the activity of a yet unidentified Ser7 phosphatase (Tietjen et al., 2010). Another pair of Ser2 elongation kinases is also present in the fission yeast S*chizzosaccharomyces pombe* (*Sp)*, where Lsk1, the *Sp* orthologue of Ctk1, has been shown to be responsible for the bulk of Ser2 phosphorylation while *Sp*CDK9, the *Sp* orthologue of Bur1, is able to phosphorylate both the CTD and Spt5 (Karagiannis and Balasubramanian, 2007; Viladevall et al., 2009).

Until recently, higher eukaryotes appeared to have only one Ser2 CTD kinase: P-TEFb (which is composed of CDK9 and cyclinT). P-TEFb is able to phosphorylate both the Ser2 position of the CTD and the elongation factor Spt5 and is essential for transcriptional elongation, as it overcomes transcriptional pausing: RNAPII is "paused" near the promoter by the activity of negative transcription elongation factors (DSIF and NELF) which trap the elongation complex in an arrested state, P-TEFb is required to overcome the influence of these negative elongation factors (Marshall et al., 1996; Marshall and Price, 1992; Marshall and Price, 1995)[6].

[6] For detailed discussions see (Lenasi and Barboric, 2010), (Peterlin and Price, 2006), and (Nechaev and Adelman, 2011)

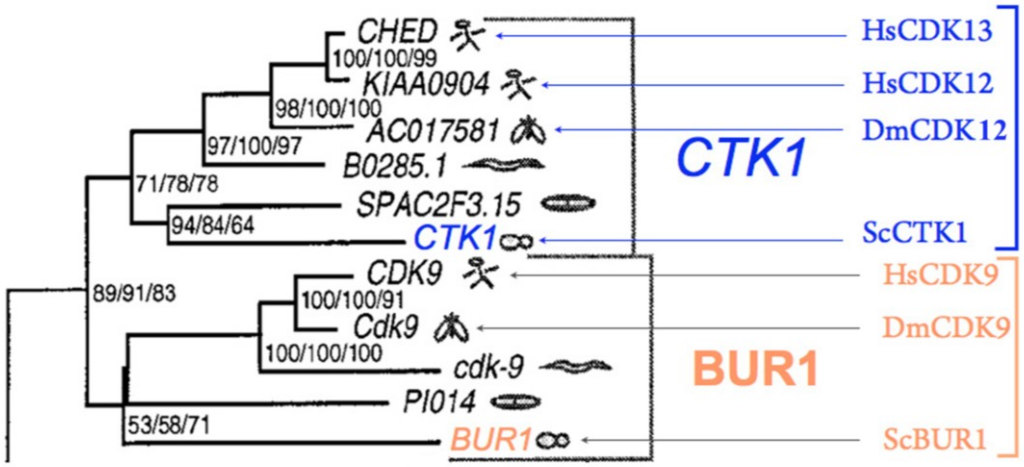

Figure 1.3: CDK evolutionary relationships as determined by Liu and Kipreos (Liu and Kipreos, 2000) using the neighbor-joining method (NJ). Bootstrap support values are given at branch points and are derived from maximum-likelihood/NJ/gamma-rate-corrected NJ analyses. Modified from (Liu and Kipreos, 2000) by Dr. Arno Greenleaf (Bartkowiak et al., 2010).

The substrate specificity of P-TEFb, coupled with its equal sequence similarity to both Bur1 and Ctk1, has led to the proposal that P-TEFb reconstitutes the activities of both yeast kinases in higher eukaryotes (Wood and Shilatifard, 2006). However, there was some evidence that this may not be the case; two evolutionary studies concluded that while *Sc* Bur1 is the closest *Sc* relative of metazoan CDK9 proteins, *Sc* Ctk1 is actually more closely related to another set of relatively little studied metazoan CDK proteins (Guo and Stiller, 2004; Liu and Kipreos, 2000) (Fig 1.3). Therefore the previously unstudied *Drosophila* CDK12 (dCDK12) and the little-studied human CDK12 (hCDK12) and human CDK13 (hCDK13) have surfaced as the potential missing partners to P-TEFb, conceivably extending the paradigm of Ser2 CTD kinase pairs to higher eukaryotes (Fig. 1.4).

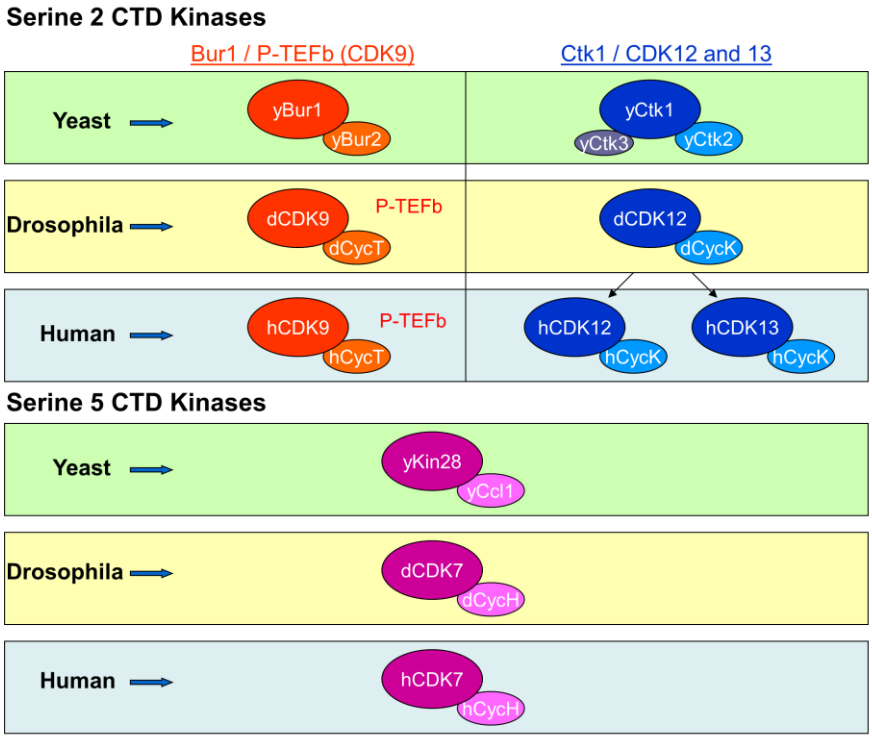

Figure 1.4: A summary of the kinase and cyclin homologs of the CTD kinases in the yeast *Saccharomyces cerevisiae*, *Drosophila*, and human. CDKs are shown as large ovals, cyclin partners are illustrated by smaller ovals which are offset to the lower right; accessory factors are depicted like cyclins, but are offset to the lower left.

1.4.3 CDK12 and CDK13

The hCDK12 and hCDK13 kinases were discovered serendipitously by probing of cDNA libraries for cholinesterase and CDK related sequences (Ko et al., 2001; Lapidot-Lifson et al., 1992; Marques et al., 2000). Interestingly, hCDK12 and hCDK13 are uncharacteristically large for CDKs with molecular weights of 160 and 180 kDa

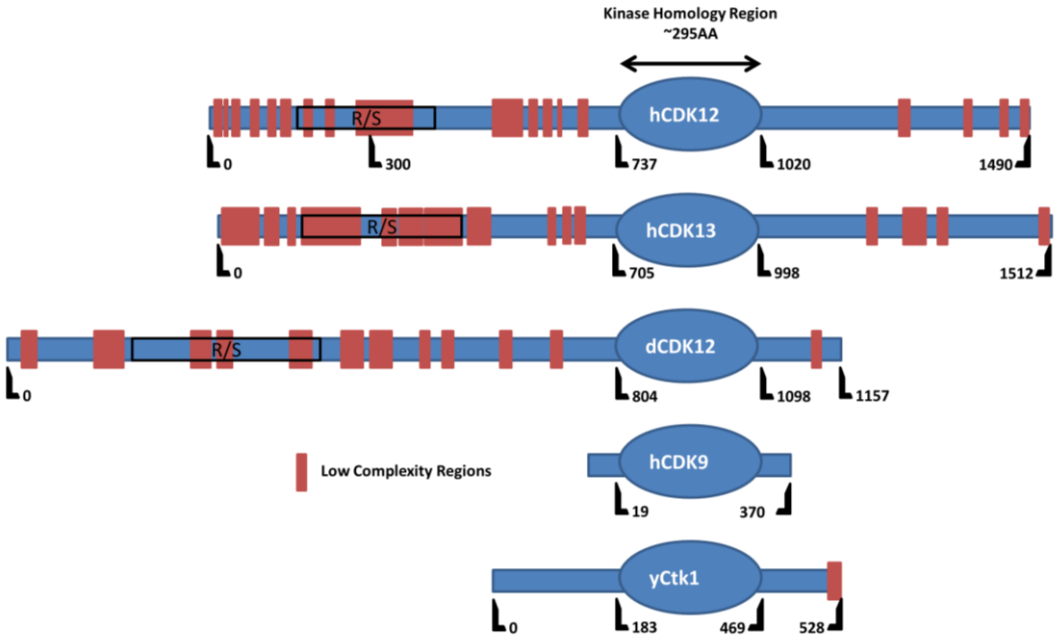

Figure 1.5: Schematics of hCDK12, hCDK13, dCDK12, hCDK9, and yCtk1 primary structures. Relative amino acid positions are indicated below each sequence with the kinase homology domain highlighted as an oval. Unlike most CDKs, CDK12 and CDK13 contain long N and C terminal arms that extend from the Ser/Thr kinase homology domain. These arms contain stretches of low complexity sequence (marked in red) and RS domains; motifs characteristic of splicing factors and believed to mediate protein-protein interactions.

respectively (most CDKs are about 50 kDa), and contain large N and C terminal extensions that flank the CDK kinase domain (Even et al., 2006; Ko et al., 2001). These extensions contain multiple stretches of low complexity sequence and RS domains; domains characteristic of splicing factors and believed to mediate protein-protein interactions (Zhong et al., 2009) (Fig 1.5). Although hCDK12 had been shown to have an *in vitro* CTD kinase activity in an early study (Ko et al., 2001), subsequent research focused on the role of hCDK12 and 13 in splicing. As expected of RS proteins, hCDK12 and hCDK13 were found to localize in nuclear speckles (believed to be storage sites for splicing factors (Zhong et al., 2009)) and their overexpression or knockdown was shown

to alter the splicing of the E1A reporter minigene *in vivo* (Berro et al., 2008; Chen et al., 2006; Chen et al., 2007). The mechanism for this splice site modulation was largely unexplored, as was any potential role of the kinases in transcription. Based on the previously discussed phylogenetic data, the aim of this dissertation was to determine whether the previously unstudied dCDK12 and little-studied hCDK12 and hCDK13 were elongation phase CTD kinases, the metazoan orthologues of yeast Ctk1, and to characterize their functions *in vivo* and *in vitro*.

2. Characterization of *Drosophila* and Human CDK12 as Elongation Phase RNA Polymerase II C-terminal Domain Kinases[1]

2.1 Introduction

The C-terminal repeat domain (CTD) of RNA polymerase II (RNAPII) is an intrinsically unstructured extension of the enzyme's largest subunit, Rpb1. The CTD is composed of a tandem array of seven amino acid repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$; the number of heptad repeats varies from organism to organism and appears to correlate with genomic complexity (Corden, 1990). Although dispensable for polymerase activity *in vitro*, the CTD is conserved throughout evolution and is essential to life (Zehring et al., 1988). The CTD's function is to coordinate transcription with nuclear processes such as mRNA processing and chromatin modification; it fulfills this role by serving as a selective binding scaffold for nuclear factors[2]. The binding specificity of the CTD for particular factors is determined by its phosphorylation state, which varies during the transcription cycle; Ser2, Ser5, and Ser7 of the heptad repeat have been identified as the primary targets of phosphorylation, with Tyr1 and Thr4 phosphorylation being recently recognized as functionally relevant (Chapman et al., 2007; Egloff et al., 2007; Eick and Geyer, 2013; Hintermair et al., 2012; Hsin et al., 2014;

---

[1] Adapted from, and published in, Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., Greenleaf, A. L., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24**,** 2303-16.

[2] See (Bartkowiak et al., 2011; Corden, 2013; Phatnani and Greenleaf, 2006) for reviews.

Hsin et al., 2011; Mayer et al., 2012). Initiating and early-transcribing polymerases are thought to be phosphorylated at Ser5 positions in the CTD repeats, whereas actively-elongating polymerases carry phosphates on both Ser2 & Ser5 residues. Toward the 3' end of pre-mRNA transcription units Ser2P appears to predominate (Eick and Geyer, 2013; Phatnani and Greenleaf, 2006).

The phosphorylation of CTD Ser5 residues during initiation and early transcription is mediated by the TFIIH CTD kinase subunit: Kin28 in yeast and CDK7 in metazoans (Komarnitsky et al., 2000; Schroeder et al., 2000). The majority of Ser2 phosphorylation on productively elongating RNAPII in *Saccharomyces cerevisiae* (*Sc*) appears to be catalyzed by CTDK-I (Buratowski, 2009; Qiu et al., 2009), whose kinase subunit is Ctk1 (Lee and Greenleaf, 1991; Sterner et al., 1995). Although it is responsible for the bulk of Ser2 phosphorylation *in vivo*, Ctk1 enzyme is not the only elongation-phase kinase in yeast; it coexists with the essential Bur1 kinase (Yao et al., 2000). It has been proposed that Bur1 kinase phosphorylates the transcription elongation factor subunit Spt5, rather than the CTD of Rpb1 (Zhou et al. 2009). However, the evidence suggests that the situation is more complex and that, in addition to acting on Spt5, Bur1 phosphorylates Ser2 positions in the CTD (Qiu et al., 2009). A similar pair of Ser2 CTD kinases is also present in the fission yeast S*chizzosaccharomyces pombe (Sp)*: Lsk1, the Ctk1 ortholog in *Sp*, phosphorylates Ser2 positions in the CTD (Coudreuse et al., 2010;

Karagiannis and Balasubramanian, 2007), whereas *Sp*CDK9, the Bur1 homologue phosphorylates both Rpb1 (on its CTD) and Spt5 (Viladevall et al., 2009).

With respect to the principal Ser2 kinases in metazoa, the situation has been less clear cut. It was previously thought that there is only one Ser2 CTD kinase in higher organisms called P-TEFb (consisting of CDK9 and a cyclin, usually cyclinT but also cyclinK) (Fu et al., 1999; Peterlin and Price, 2006). P-TEFb is able to phosphorylate both the CTD and the elongation factor subunit Spt5 (Kim and Sharp, 2001). The dual functionality of CDK9, coupled with its equal sequence similarity to both Bur1 and Ctk1, has led to the assumption that CDK9 reconstitutes the activities of both Bur1 and Ctk1 in higher eukaryotes (Wood and Shilatifard, 2006). On the other hand, two independent molecular evolution studies concluded that *Sc* Bur1 is the closest *Sc* relative of metazoan CDK9, while *Sc* Ctk1 is closer to a set of largely unstudied metazoan CDKs (Guo and Stiller, 2004; Liu and Kipreos, 2000), the "Ctk1" group (Fig. 1.3). The genomes of *Drosophila melanogaster (Dm)* and *Caenorhabditis elegans* each encode one member of the Ctk1 group (genes CG7597, and B0285.1, respectively), while the genome of *Homo sapiens* encodes two (genes CRKRS and CDC2L5; see Fig 1.3). The human "Ctk1" proteins have been renamed CDK12 and CDK13 (Chen et al., 2006; Chen et al., 2007). We will refer to them as hCDK12 and hCDK13 and to the *Dm* protein as dCDK12.

The evolutionary studies thus suggest that the metazoan ortholog of Ctk1 is dCDK12 in *Drosophila* and the isozyme pair, hCDK12 and hCDK13 in humans. Both the

evolutionary arguments and the recent experimental work mentioned above suggest that the metazoan ortholog of Bur1 is CDK9. We tested these suggestions with a set of diverse experiments which lead us to conclude that, indeed, the human and *Drosophila* orthologs of yeast Ctk1 are CDK12 and CDK13. Because the metazoan orthologs of yeast Ctk1 have not previously been studied as CTD kinases, these experiments open the door to investigating a vast array of transcription-related events in metazoa that are critical to many aspects of proper gene expression and other nuclear functions.

2.2 Materials and methods

2.2.1 Antibodies and western blotting

Anti –IIo: Goat affinity-purified IgGs directed toward the hyperphosphorylated CTD of *Dm* Rpb1 (Morris et al., 1997; Weeks et al., 1993). rAP anti-dCDK12: Rabbit affinity-purified IgGs directed against a peptide comprising amino acids 705–722 of dCDK12 (CG7597; NP_649325.2). rAP anti-hCDK13: Rabbit affinity-purified IgGs directed against a peptide comprising amino acids 294–312 of hCDK13 (CDC2L5. NCBI RefSeq: NP_003709.3). rAP anti-hCDK12: Rabbit affinity-purified IgGs directed against a peptide comprising amino acids 1058–1077 of hCDK12 (CRKRS. NCBI RefSeq: NP_057591.2). Anti-dCyclin T: Affinity-purified sheep IgG against purified dCyclin T. Anti-NonP: Mouse monoclonal 8WG16 (Abcam). Anti-Ser2P: Rat monoclonal (3E10) from D.Eick (Chapman et al., 2007), or mouse monoclonal (H5) from Covance. Anti-Ser5P: mouse monoclonal (H14) from Covance. Antibody to b-actin: A mouse

monoclonal from Abcam. gAP-anti-dRpb2: Affinity-purified goat IgG against a segment from dRpb2 (amino acids 519–992) (Skantar and Greenleaf, 1995). Western blot analysis was performed using the Odyssey infrared scanner and secondary antibodies from Li-Cor.

2.2.2 Polytene immunofluorescence

Visualization of proteins on fixed polytene chromosomes (Weeks et al., 1993) was carried out as described previously (Schwartz et al., 2004). Briefly, *Drosophila* were grown in bottles with yeast glucose media at 23 °C and wandering third instar larvae were collected. For heat shock, larvae were wrapped in plastic and submerged in a 36.5 °C water bath for the indicated amount of time. Larvae were dissected in 100 µL of dissection buffer (0.7% NaCl) and the salivary glands were isolated and transferred to 40 µL of fix buffer (50 µl 37% paraformaldehyde, 450 µl acetic acid, 500 µl H_2O) on a microscope slide. After 10-20 min of incubation the salivary glands were squashed using a coverslip and visualized using a microscope; well-spread and clearly banded chromosomes were frozen in liquid nitrogen, and then stained.

For staining, slides were removed from liquid nitrogen, coverslips were removed, and slides were washed (PBS and then PBS with 1% Triton-X 100). Blocking was performed using PBS with 5% non-fat milk powder for 60 min and primary antibodies were added in PBS and 1% BSA overnight at 4 °C. Slides were then washed and secondary antibody solution was added (antibodies in PBS, 1% BSA, and 2%

donkey or goat serum) for 1 hour at room temp; secondary antibodies were from Jackson ImmunoResearch - conjugated with RhodamineRedX and Cy2. Slides were washed using 300 and then 400 mM NaCl, 0.2% NP40, 0.2% Tween20, stained with DAPI, and imaged using a Zeiss Axioplan Microscope with a 25x or 100x lens; images were processed with Openlab software.

2.2.3 Chromatin Immunoprecipitation

Drosophila S2 cells were grown in M3 media + BPYE (Bacto-peptone, yeast extract) with 10% FBS to a density of $6x10^6$ to $7x10^6$ cells per mL. To prepare the 10 min heat-shock chromatin, 4 mL of 48 °C M3 + BPYE (no serum) was added to 4 mL of culture and incubated for 10 min at 36.5 °C, and then 4 mL of 4 °C M3 + BPYE (no serum) was added to bring the cells to room temperature. Immediately, formaldehyde was added to a final concentration of 1% and mixed for 2 min at room temperature. The reaction was quenched with 675 µL of 2.5 M glycine and immediately cooled on ice. The cells were centrifuged for 5 min at 4 °C and resuspended in sonication buffer (20 mM Tris-Cl at pH 8.0, 2 mM EDTA, 0.5 mM EGTA, 0.5% SDS, 0.5 mM PMSF, Roche complete protease inhibitor cocktail [catalog no. 05 056 489 001]) at 10^8 cells per milliliter. The cells were sonicated 12 times for 20 sec each, with a 1 min rest in between at 4 °C using a Bioruptor sonicator; the material was centrifuged at 20,000xg for 10 min at 4 °C, and 25 µL of the supernatant was used per immunoprecipitation. Pol II was immunoprecipitated with 4 µL of rabbit anti-Rpb3 antisera (Lis laboratory stock) and

dCDK12 was immunoprecipitated with 7 µL of rAP anti-dCDK12. The non-heat-shock chromatin was prepared in the same manner, except 8 mL of room temperature M3 + BPYE (no serum) was added to 4 mL of culture. The amount of immunoprecipitated DNA was quantified using a Roche LightCycler 480 by comparing the quantification cycle of the immunoprecipitated DNA to the quantification cycle of serial dilutions of input DNA to obtain the percent input.

2.2.4 Immunoprecipitation and purification of hCDK13 and dCDK12

2.2.4.1 dCDK12 purification

2 mL of Kc cell nuclear extract (Price et al., 1987) were diluted to a final volume of 4 mL in ice-cold HGEDPF (25 mM HEPES, 8% glycerol, 0.1 mM EDTA, 1 mM DTT, phenylmethylsulfonyl fluoride [1:1000 dilution of saturated solution in isopropanol], 10 mM NaF); NaCl and TritonX-100 were added to a final concentration of 0.5 M NaCl and 0.4% Triton X-100. The extract was then centrifuged at 23,000xg for 1 hour at 4 °C, and the supernatant was precleared by rocking end-over-end with 20 µL of protein A/G bead slurry (Thermo Scientific) for 1 h at 4 °C. Approximately 50 µL of protein A/G beads (Thermo Scientific, Ultra-Link resin) were saturated with rAP anti-dCDK12 IgG and added to the precleared extract, and then allowed to rock end-over-end overnight at 4 °C. The beads were recovered, transferred to a small disposable column (Bio-Rad), and washed at 4 °C with 10 mL of PBS + 0.2% Tween 20, then with 10 mL of HGEDPF at 0.5 M NaCl, followed by another 10 mL of PBS + 0.2% Tween20. The beads were then

transferred to a fresh microfuge tube and resuspended in 100 mM NaCl/HGEDPF. Beads used directly for kinase assays were further washed in 0.8 M NaCl/HGEDPF before being rinsed and resuspended in 1x kinase buffer (25 mM HEPES at pH 7.6, 10 mM $MgCl_2$, 5 mM NaF).

For dCDK12 protein elution, protein A acrylic beads (Sigma, catalog no. P1052) were saturated with anti-dCDK12 antibody and 100 µL of beads were incubated with 2.5 mL of KcN extract for 2 hours at 4 °C with end-over-end rocking in a Bio-Rad disposable column. The beads were washed in HGEDP (same as HGEDPF, but without NaF) with 0.5% NP40 and 100 mM KCl, followed by 400 mM KCl/HGEDP and 100 mM KCL/HGEDP without detergent. Elution was carried out by an 1 hour incubation with end-over-end rocking at 4 °C with 1 mM antigenic peptide (ABR/Thermo Scientific) in 100 µL of 100 mM KCl/HGEDP. Elutions 2 and 3 were performed in the same manner, but with a concentration of 100 mM antigenic peptide. For assays, 5 µL of each elution was used.

2.2.4.2 hCDK13 purification

HeLa nuclei were prepared, extracted with 0.15 M NaCl, and centrifuged (Longley et al., 1997). 25 mL of the pellet was used to prepare a 0.5 M NaCl extract of the nuclear pellet as described previously (Carty et al., 2000). The 0.5 M NaCl extract was filtered through Miracloth (Chicopee Mills, Inc.) to remove floating lipid, and protein concentration was determined using the Bradford method (Bio-Rad Protein

Assay). Ion exchange chromatography was performed on a 25 mL Macro-Prep CM (Bio-Rad) column. Extract containing 400 mg of protein was diluted to 0.15 M NaCl with HGEDPF and batch-bound to the equilibrated Macro-Prep CM slurry for 2 hours at 4 °C. The slurry was then loaded back into the column, washed with 0.15 M NaCl/HGEDPF, and eluted using steps of 0.25, 0.4, 0.6, and 1 M NaCl in HGEDPF. The bulk of hCDK13 was found to elute in the 0.6 M NaCl fraction (monitored by Western blot). The 0.6 M fractions containing hCDK13 were pooled and frozen at -80 °C in 1 mL aliquots. Antibody-saturated beads were prepared with Ultra-Link protein A/G beads (10–100 µL of slurry) or magnetic protein G Dynabeads (50 µL of slurry) and rAP anti-hCDK13 were mixed with the 0.6 M NaCl CM column fractions (1–8 mL) and allowed to rock end-over-end for 8 hours to overnight. The beads were then washed with multiple bead bed volumes of each of the following: PBS/0.2% Tween-20, 0.8 M NaCl/HGEDPF, PBS/0.2% Tween20, and, finally, 1x kinase reaction buffer (25 mM HEPES at pH 7.6, 10 mM $MgCl_2$, 5 mM NaF). The beads were resuspended in 1x kinase reaction buffer, frozen in liquid nitrogen, and stored at -80 °C.

2.2.5 CTD kinase assays

CTD kinase assays were performed essentially as described (Jones et al., 2004; Morris et al., 1997), with minor variations. Briefly, dCDK12 bound to Protein A/G UltraLink beads (Thermo Scientific) and hCDK13 bound to magnetic Protein G Dynabeads (Invitrogen) were used to assay for *in vitro* CTD kinase activity (see kinase

purification). 5 μL of Dynabeads (from 50 μL of antibody saturated dynabead slurry incubated with 1 mL of pooled hCDK13 CM column fractions) or Ultralink (from 100 μL of antibody saturated protein A/G bead slurry incubated with 2 mL of KCN extract) slurry was used for the kinase assay. Each reaction was performed in a final concentration of 25 mM HEPES pH 7.6, 10 mM $MgCl_2$, 5 mM NaF, and 60 μM ATP, in the presence of [γ-^{32}P]ATP (1 μL at 10 μCi), with or without ~2 μg GST-yCTD substrate, in a total volume of 10 μL. The reactions were allowed to proceed for 1 hour at 30 °C, and were stopped by the addition of 10 μL 2xSDS and boiling for 10 min at 95 °C. The reactions were then run on a 10% SDS-PAGE gel (Biorad). The gel was stained with coomassie and exposed to a storage phosphor screen (Molecular Dynamics) for 15 min.

2.2.6 RNAi depletion of dCDK12 in S2 cell culture

dsRNA constructs targeting CG7597 were designed using the SnapDragon program (http://www.flyrnai.org/cgi-bin/RNAi_find_primers.pl), which is optimized for the design of long dsRNAs for *Dm*. Three constructs were selected (targeting exon 2, exon 4, and exon 7 of CG7597), and the primers were extended with the sequence of the T7 promoter. Constructs used were as follows:

E2F, 5'-GAATTAATACGACTCACTATAGGGAATCAATGCCTTACTGGACCG-3';

E2R, 5'-GAATTAATACGACTCACTATAGGGATGGTATCGTGACTGCTCGAC-3';

E4F, 5'-GAATTAATACGACTCACTATAGGGAGTGTGCAGCACTCACGAAGT-3';

E4R, 5'-GAATTAATACGACTCACTATAGGGACTCTTGCCGCTCAATGATTT-3';

E7F, 5′-GAATTAATACGACTCACTATAGGGACTCCTACGGTGATAGCCTCG-3′;

E7R, 5′-GAATTAATACGACTCACTATAGGGAATCTGATGCTGTGTGCGTTC-3′.

The primers were used to amplify DNA from Drosophila genomic DNA (a gift from Dr. Tao Hsieh's laboratory). dsRNA was made using a MEGAScript T7 Kit (Ambion, Inc.) following the manufacturer's instructions; the resulting RNA was purified using phenol chloroform extraction and annealed. S2 cells were grown in M3 + BPYE with 10% serum. Cells (1.5 mL of $1x10^6$ cells per milliliter) were silenced by addition of 15 μg of dsRNA in a sixwell plate format for 48 hours. Anti-LacZ dsRNA (Adelman lab) was used as a control. After 48 hours, cells were resuspended, pelleted, washed in PBS, repelleted, and lysed using 2x SDS loading buffer.

2.2.7 RNAi depletion of hCDK12 and hCDK13 in human cell culture

The HeLa R-19 cell line was a gift from Dr. Shelton Bradrick. HeLa R-19 was cultured in DMEM with 10% FBS and 1x PenStrep (all from GIBCO). Experimentally verified FlexiTube siRNAs targeting hCDK13 (Hs_CDC2L5_5 and Hs_CDC2L5_6) and hCDK12 (Hs_CRK7_5 and Hs_CRK7_6) were purchased from Qiagen. The control siRNA was AllStarsNegative Control siRNA from Qiagen. HeLa R-19 cells were reverse-transfected in six-well plates and/or 60-mm dishes using RNAi Max (Invitrogen) transfection reagent following the manufacturer's instructions. One siRNA targeting hCDK12 (Hs_CRK7_5) was found to have an off-target effect that resulted in cell death, and was not used. The second hCDK12 siRNA and a pool of the two hCDK13 siRNAs

were used for the knockdown experiments. Knockdown experiments were conducted with seeding densities calculated so that the cells reach ~90% confluence before being harvested. Final siRNA concentrations of 5 nM and 12.5 nM (total concentrations for both the pooled [CDK13] and single [CDK12] siRNAs) and time points of 48 and 72 hours were found to yield the same results. Cells were harvested by trypsinization with 0.05% Trypsin-EDTA (GIBCO) for 3 min at 37 °C, followed by neutralization of the trypsin with complete medium. The resulting cell slurry was centrifuged at low speed, and the cell pellet was washed by resuspension in 5 mL of PBS. The cells were pelleted, resuspended in 1 mL of PBS, and repelleted. After removing the supernatant, the cell pellet was lysed with RIPA buffer or hot 2x SDS loading buffer (final volume ~130 μL). Successful knockdown of the kinases was confirmed by Western blot analysis.

2.3 Results

2.3.1 dCDK12 colocalizes with actively transcribed genes on *Drosophila* polytene chromosomes

In order to visualize the genome-wide distribution of dCDK12 on chromatin, we began our studies with indirect immunofluorescence labeling of *Drosophila* polytene chromosomes. According to our hypothesis and previous experiments examining the distribution of yeast Ctk1 (Cho et al., 2001; Qiu et al., 2009), we expected to find dCDK12 localized across actively transcribing loci in parallel with hyperphosphorylated RNAPII (Pol II0, a marker that correlates with actively elongating polymerase). In collaboration with Dr. John Lis's lab, the following immunoflourescence experiments were performed

by Janis Werner. Salivary glands were dissected from *Drosophila* larvae, with and without prior heat shock[3], fixed with formaldehyde, and squashed on a microscope slide. The spread polytene chromosomes were then stained with antibodies against dCDK12 and Pol II0 (Weeks et al., 1993). dCDK12 colocalized extensively with hyperphosphorylated RNAPII in both heat shocked (Fig. 2.1 a.) and non-heat shocked animals (Fig. 2.1 b.). Split chromosome arm views further demonstrate the extent of this colocalization (Fig. 2.1 c. and d.); practically all elongating RNAPII is accompanied by

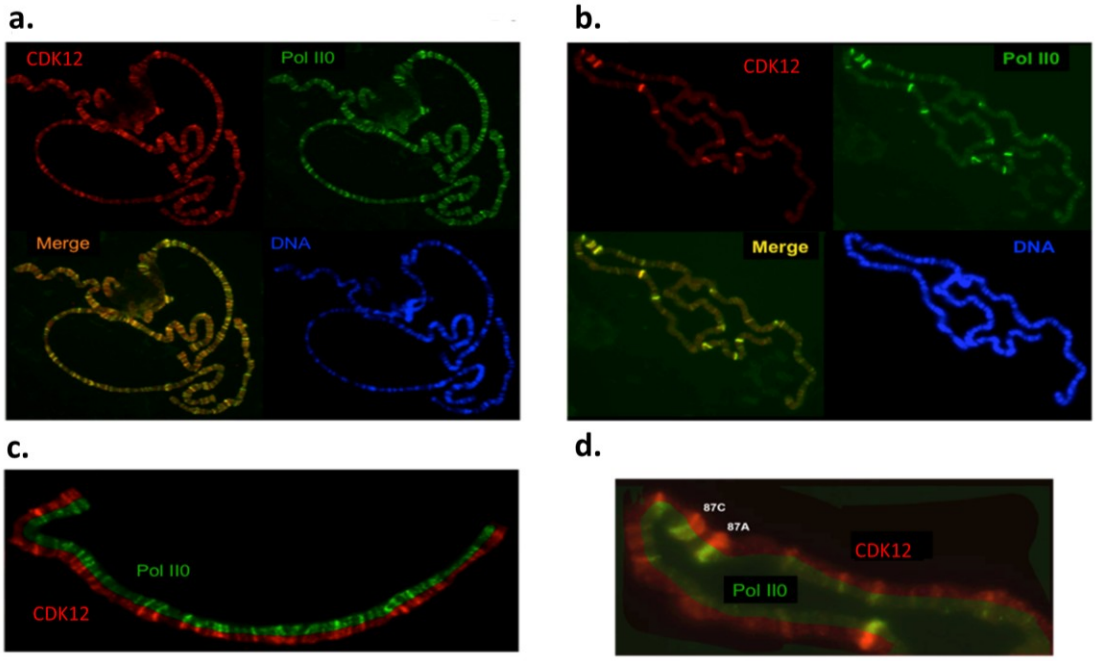

Figure 2.1: Distribution of dCDK12 (red) and hyper-phosphorylated RNAPII (Pol II0; green) on *Drosophila melanogaster* polytene chromosomes from (a.) non-heat shocked animals and (b.) heat shocked animals; DNA is stained with DAPI (blue) and the merge contains both the green and red images. Split chromosome arm views from (c.) non-heat shocked and (d.) heat shocked animals; the 87A and 87C heat shock puffs are labeled. Images created by Dr. Arno Greenleaf.

[3] 20 min at 37 °C; heat shock results in the suppression of the majority of transcription, with the exception of the heat shock genes which are strongly upregulated.

dCDK12. This indicates that dCDK12 is present at sites of active transcription and suggests that it is likely to play a role during transcriptional elongation.

2.3.2 The global distributions of dCDK12 and P-TEFb on *Drosophila* polytene chromosomes are distinct

In an effort to begin to deconvolute the relationship between P-TEFb and CDK12 we also compared the genome-wide distributions of dCDK12 and P-TEFb. Due to the lack of an antibody against CDK9 itself, we obtained antibodies against dCyclinT, the cyclin subunit of P-TEFb (Cdk9 + cyclinT) from Dr. David Price's lab. dCDK12 and P-TEFb exhibited very different distribution patterns on both heat shocked and non-heat shocked chromosomes (Fig. 2.2). Interestingly the long, intron rich, ecdysone induced developmental genes stained very strongly for dCDK12, and weakly for dCyclinT. Other regions of the chromosome showed the inverse relationship, and although clearly distinct, the distributions of the two kinases did share some areas of overlap (see merge in Fig. 2.2 a. and Fig 2.2 d.). It is worth noting that the kinase distributions can change drastically across a single transcription puff. In (Fig 2.2 d.) the heat shock puff, located between the two dotted lines, appears to be flanked by a very strong dCyclinT signal (green) with virtually no signal from dCDK12 (red). Moving further up the puff, the strong dCyclinT signal then intersects with the appearance of a robust dCDK12 signal, creating an area of overlap (yellow band in the merged image). The dCyclinT signal

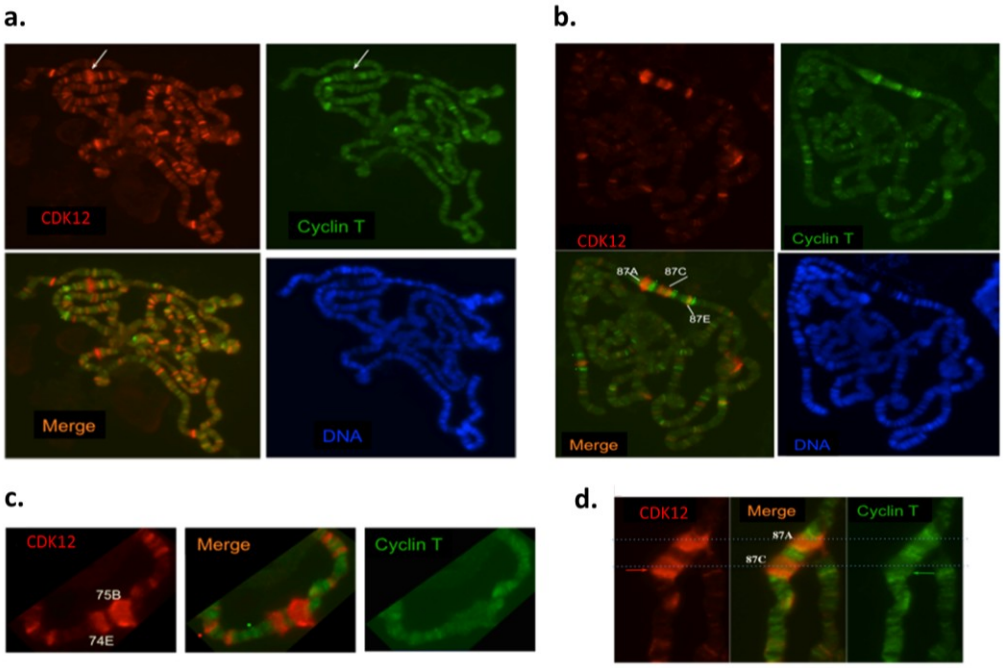

Figure 2.2: Distribution of dCDK12 (red) and P-TEFb (CDK9+CyclinT; green) on *Drosophila melanogaster* polytene chromosomes from (a.) non-heat shocked animals and (b.) heat shocked animals; DNA is stained with DAPI (blue) and the merge contains both the green and red images, white arrows indicate developmental puffs that stain strongly for dCDK12, but weakly for P-TEFb. Chromosome arm views from (c.) non-heat shocked (the 75B and 74E ecdysone developmental puffs are labeled) and (d.) heat shocked animals (the 87A, 87C, and 87E heat shock puffs are labeled). Images created by Dr. Arno Greenleaf.

then drops off, and gives way to the strong dCDK12 signal (red, near top dotted line), which dominates through the rest of the puff. These results strongly support the idea that dP-TEFb and dCDK12 have distinct functions at specific locations within the transcription unit.

2.3.3 The distribution of dCDK12 across the Hsp70 loci is distinct from P-TEFb

In order to verify the polytene immunofluorescence results and to determine the distribution of dCDK12 at higher resolution we set out to perform dCDK12 and RNAPII

ChIP at the heat shock inducible Hsp70 loci. The following set of ChIP experiments was performed in collaboration with Dr. John Lis's lab by Nicholas Fuda. Using *Drosophila* S2 cell culture, we find that on the well-studied Hsp70 gene, anti-Rpb3 antibody detects the expected presence of promoter-proximally paused polymerase near the TSS prior to heat shock (Fig 2.3 a.; +96 position, blue bars). After a 10 min heat shock, RNAPII occupancy increased dramatically along the entire transcription unit due to RNAPII recruitment and its release into productive elongation; however, it is important to note that a peak of RNAPII occupancy persists at the +96 position as pause release remains a rate limiting step (Fig 2.3 a.; red bars). As opposed to RNAPII ChIP, dCDK12 ChIP revealed no signal above background before heat shock, but displayed a robust signal along virtually the entire gene after heat shock (Fig 2.3 b.). Significantly, the results indicate no presence of dCDK12 with the paused RNAPII before gene activation, and no dCDK12 peak at the +96 position following gene activation. Maximum levels of dCDK12 are not reached until the +379 position and remain high throughout the transcription unit (Fig 2.3 b.). These results contrast with those for P-TEFb, which reveal a concentration that peaks at the 5' end of the gene (at the +58 position in (Boehm et al., 2003)) and steadily declines, as would be expected due to P-TEFb's role in pause release. The ChIP results, therefore, are consistent with our polytene chromosome-staining results in showing different distributions for dCDK12 and dP-TEFb and reveal a dCDK12 occupancy pattern reminiscent of an elongation phase Ser2 CTD kinase.

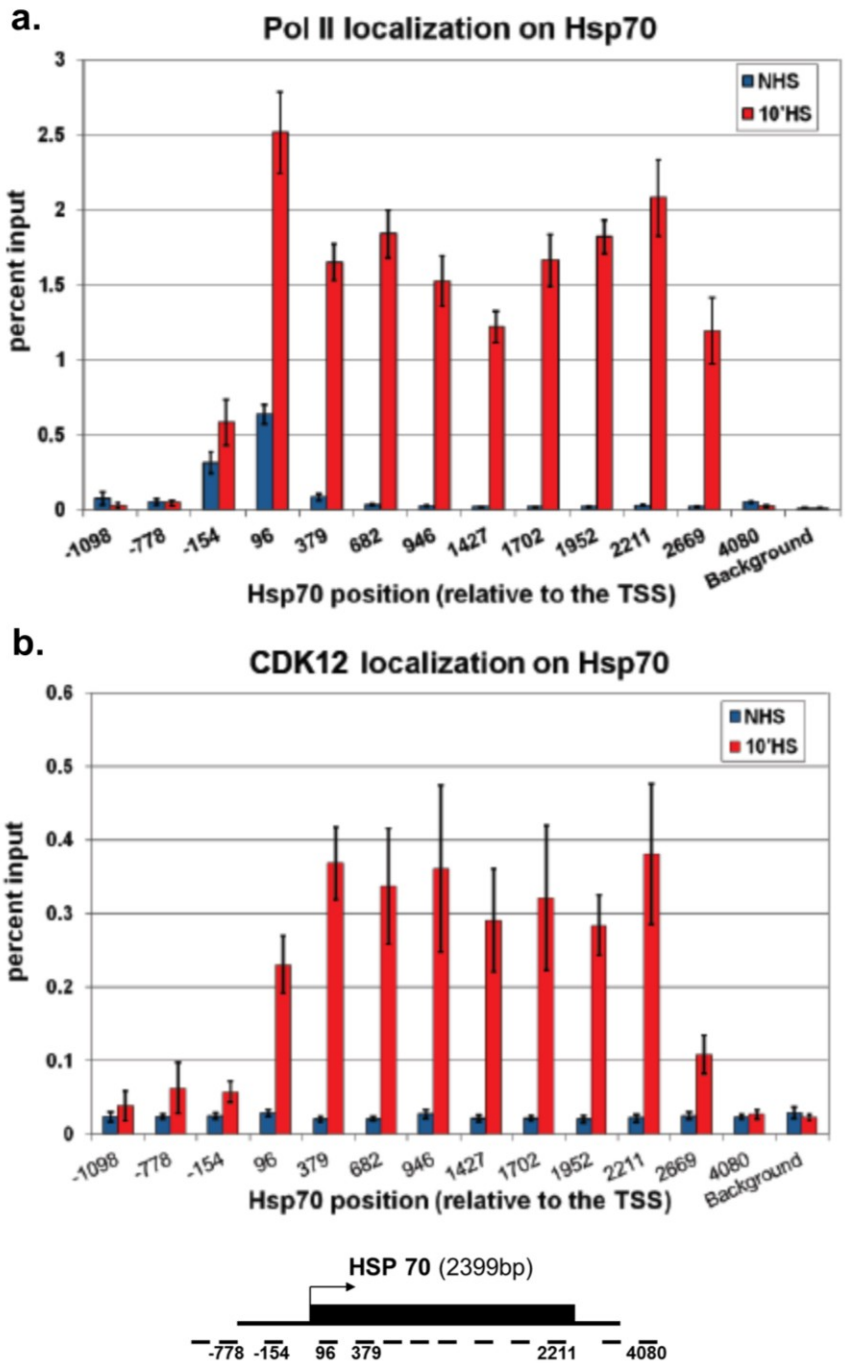

Figure 2.3: ChIP analysis of (a.) RNAPII and (b.) dCDK12 at the Hsp70 loci. Cells were either non-heat shocked (in blue) or heat shocked for 10 minutes (red). A schematic of the HSP70 gene with the positions of the assessed regions is included at the bottom of the figure. Error bars are standard error of the mean, n=3.

2.3.4 *Drosophila* CDK12 and human CDK13 manifest CTD kinase activity *in vitro*

Having investigated the distribution of dCDK12 on chromatin, we next wanted to see whether CDK12 and CDK13 were capable of phosphorylating the CTD. In order to obtain purified endogenous hCDK13 and dCDK12 for *in vitro* CTD kinase assays we employed a mixture of standard and immuno-affinity purification methods. We first turned our attention to the human proteins (hCDK12 and hCDK13), which we attempted to purify from HeLa cell nuclear pellets.

HeLa nuclear pellets were extracted in a 0.5 M NaCl extraction buffer as described in (Carty et al., 2000); the extract was then cleared and bound to a Macro-Prep CM ion exchange chromatography column. The column was eluted using a step gradient of NaCl and the bulk of hCDK13 and hCDK12 (pI 9.7 and 9.48) was found to elute in the 0.6 M NaCl fraction (monitored by western blot). In order to further purify and separate the two paralogues we employed affinity purified antibodies. Protein A/G beads were saturated with anti-hCDK13 antibodies and mixed with the 0.6 M NaCl CM column fractions for several hours. After incubation, the beads were stringently washed, resuspended in kinase reaction buffer, and assayed for CTD kinase activity (successful hCDK13 pulldown was confirmed via western blots, data not shown). Kinase assays were performed with a GST-yeastCTD (GST-yCTD) fusion protein: The yeast CTD (26 heptad repeats) fused to GST and expressed in E. Coli. ~2 ug of the purified fusion protein, 5 μL of protein A/G bead slurry, 1 μCi of [γ-^{32}P]ATP, and cold

ATP were mixed in kinase reaction buffer for a final reaction volume of 20 µL with 60 uM ATP[4]. The resulting mixture was incubated for 1 hour at 30 °C, after which the reaction was stopped with the addition of 5µL of 5x SDS loading buffer. The reaction mix was then submitted to SDS-PAGE, and the coomassie stained gel exposed to film. Reactions without fusion protein or without kinase beads were used as controls. The presence of a strong mobility shifted band (above the molecular weight of the unphosphorylated GST-yCTD protein) in the autoradiogram indicated that hCDK13 does indeed exhibit CTD kinase activity (Fig. 2.4).

Several attempts have been made to affinity purify hCDK12 from both raw nuclear extract and CM column fractions; unfortunately the hCDK12 antibody used at the time (rAP anti-hCDK12: Rabbit affinity-purified IgGs directed against a peptide

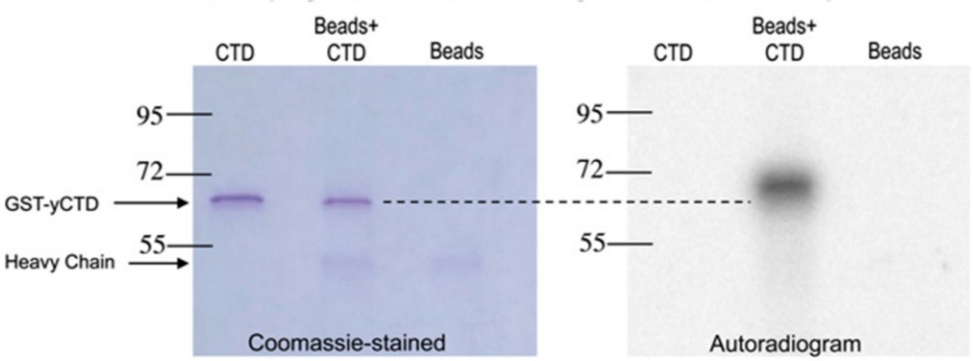

Figure 2.4: hCDK13 kinase assay. Corresponding coomassie-stained gel and autoradiograph with the position of the unphosphorylated GST-yCTD substrate indicated by a dashed line. Substrate and enzyme-bead only control lanes are labelled CTD and Beads respectively. The mobility shifted signal above the position of the GST-yCTD is indicative of CTD hyperphosphorylation.

[4] The Km of Ctk1 is ~25 uM (Jones et al., 2004).

comprising amino acids 1058–1077 of hCDK12) was unable to pull down hCDK12 in any of the conditions that were tried. This may be due to the poor specificity of this hCDK12 antibody which clearly cross-reacts with many other proteins in western blots (data not shown). Fortunately, hCDK12 CTD kinase activity has been shown previously in the literature (Ko et al., 2001), indicating that it too is capable of CTD phosphorylation *in vitro*.

dCDK12 was also purified via an immuno-affinity approach using Drosophila Kc cell nuclear extracts as starting material. The Kc nuclear extracts (Price et al., 1987) were diluted, precleared, and mixed with anti-dCDK12 antibody saturated protein A/G beads. The beads were then washed exhaustively using high salt and detergent, then resuspended in either kinase buffer or 100 mM KCl HGEDP (see 2.2.4.1 for buffer composition). Beads resuspended in kinase buffer were assayed for CTD kinase activity in the same manner as hCDK13 (Fig. 2.5 a.). In an attempt to further isolate the kinase, dCDK12 was eluted from the 100 mM KCl HGEDP resuspended beads using the dCDK12 antibody's antigenic peptide. The beads were sequentially incubated in 100 µL of HGEDP containing 1 mM (elution 1 - E1), 100 µM (elution 2 - E2), and 100 µM (elution 3 - E3) antigenic peptide, for a total time of 1 hour per elution. 5 µL of the resulting elutions were used in lieu of bead slurry for CTD kinase assays using a βgal-*Drosophila*CTD fusion protein (Fig 2.5 b.). dCDK12 exhibited *in vitro* CTD kinase activities on both GST-yCTD and βgal-dCTD substrates (Fig 2.5). Therefore, these

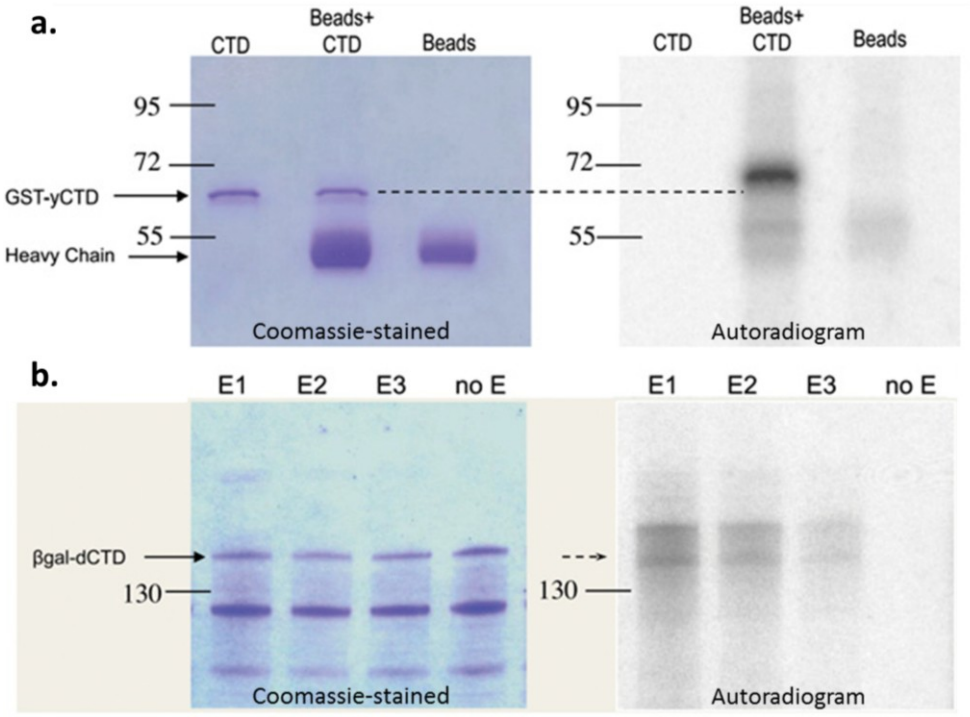

Figure 2.5: dCDK12 kinase assays. Corresponding coomassie-stained gels and autoradiographs with the position of the unphosphorylated substrate indicated by a dashed line/arrow. (a.) Bead bound dCDK12 assayed using GST-yCTD; substrate and enzyme-bead only control lanes are labelled CTD and Beads respectively. (b.) dCDK12 elutions assayed using a βgal-dCTD fusion protein. Coomassie bands below the full length substrate are degradation products. E1, E2, and E3 are subsequent elution fractions and noE is a no enzyme control lane. The mobility shifted signals above the position of the unmodified substrates are indicative of CTD hyperphosphorylation.

results clearly demonstrate that hCDK13 and dCDK12 immunoprecipitates exhibit CTD kinase activity *in vivo*.

2.3.5 *Drosophila* and human cells require CDK12 for normal phosphorylation of the CTD

In order to determine whether dCDK12 was necessary for proper CTD phosphorylation *in vivo*, we used siRNA to knockdown dCDK12 in Drosophila cell culture. Three dsRNA PCR primer pairs, designed to target exons 2, 4, and 7 of dCDK12, were amplified using PCR of Drosophila DNA and *in vitro* transcribed into

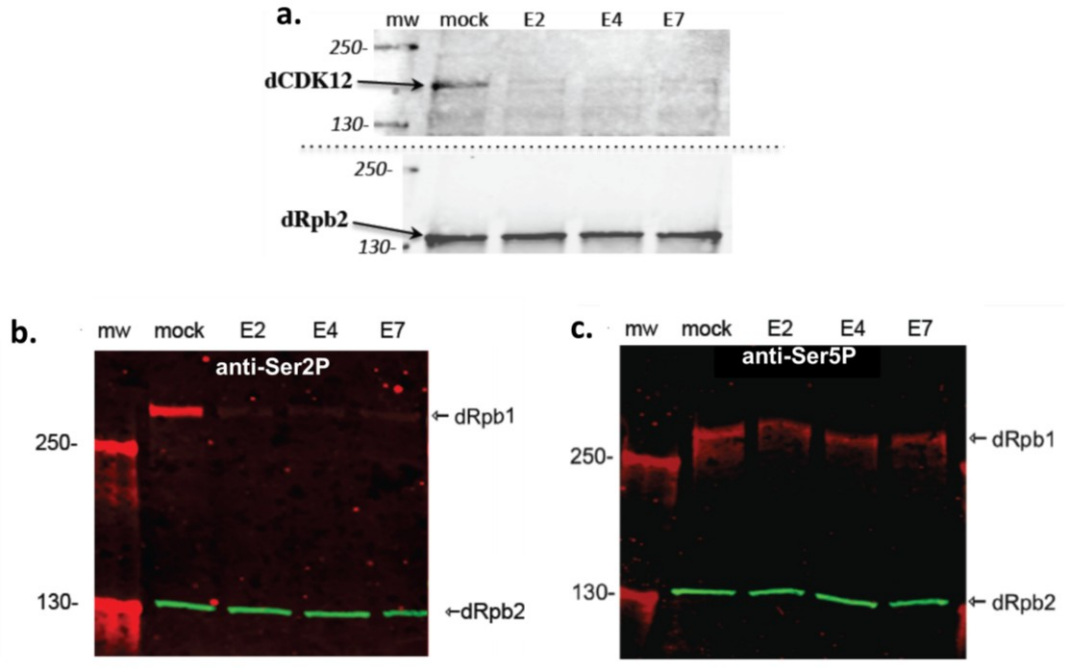

Figure 2.6: Western blots of *Drosophila* S2 whole cell extracts 48hrs post dsRNA treatment. (a.) Anti-dCDK12 western confirming dCDK12 knockdown using three dsRNA constructs (E2, E4, E7); dRpb2 is used as a loading control. (b.) Anti-Ser2P western using the H5 antibody (in red). (c.) Anti-Ser5P western using the H14 antibody (in red). Mock cells are treated with anti-LacZ dsRNA and dRpb2 (green) is used as a loading control.

dsRNA; anti lac-Z dsRNA obtained from Dr. Adelman's lab was used as a negative control (mock). The transfected cells were incubated for 48 hours and then harvested for western blotting of whole cell extracts. Westerns with anti-dCDK12 antibodies revealed that almost complete dCDK12 knockdown was obtained with all three dsRNA constructs (Fig. 2.6 a.).

In order to determine global CTD phosphorylation levels, we probed whole cell extracts with the anti-phospho-CTD antibodies H5 and H14. The H5 antibody is specific for Ser2 phosphorylated heptad repeats, while the H14 antibody reacts with Ser5 phosphorylated heptads (Jones et al., 2004). Knockdown of dCDK12 resulted in an

almost complete loss of Ser2 CTD phosphorylation as measured by the H5 antibody (Fig. 2.6 b.), and had no significant effect on Ser5 CTD phosphorylation as measured by H14 (Fig. 2.6 c.). These results clearly demonstrate that the majority of Ser2 CTD phosphorylation (as detected by the H5 antibody, see discussion) is mediated through dCDK12 and strongly support the idea that dCDK12 is a Ser2 specific CTD kinase.

In order to test the *in vivo* function of the human kinases, we also used siRNAs to knockdown hCDK12 and/or hCDK13 in HeLa cell culture. HeLa cells were reverse transfected with five different sets of siRNAs: no siRNA (RNAiMAX lipofectamine only, mock), a non-targeting siRNA pool (si-con), or siRNA pools targeting hCDK12, hCDK13, or both. The cells were then incubated for three to five days, harvested by trypsinization, and analyzed by western blotting of whole cell extracts (Fig. 2.7). Final siRNA concentrations of 5 nM and 12.5 nM, and time points of 48 and 72 hours were found to yield the same results. Western blots confirmed that we were able to successfully knockdown both hCDK12 and hCDK13, alone and in tandem. Based on our experiments in *Drosophila,* we expected to see a significant decrease in Ser2P upon the knockdown of the kinases. However, our results proved to be much more nuanced. Due to reasons previously discussed (see 1.1), the CTD containing Rpb1 subunit of RNAPII runs as two distinct bands on SDS gels: Nonphosphorylated (IIa) and hyperphosphorylated (IIo). Western blots of whole cell extracts for Ser2 phosphorylated CTD repeats probed with the 3E10 antibody (another antibody specific for Ser2

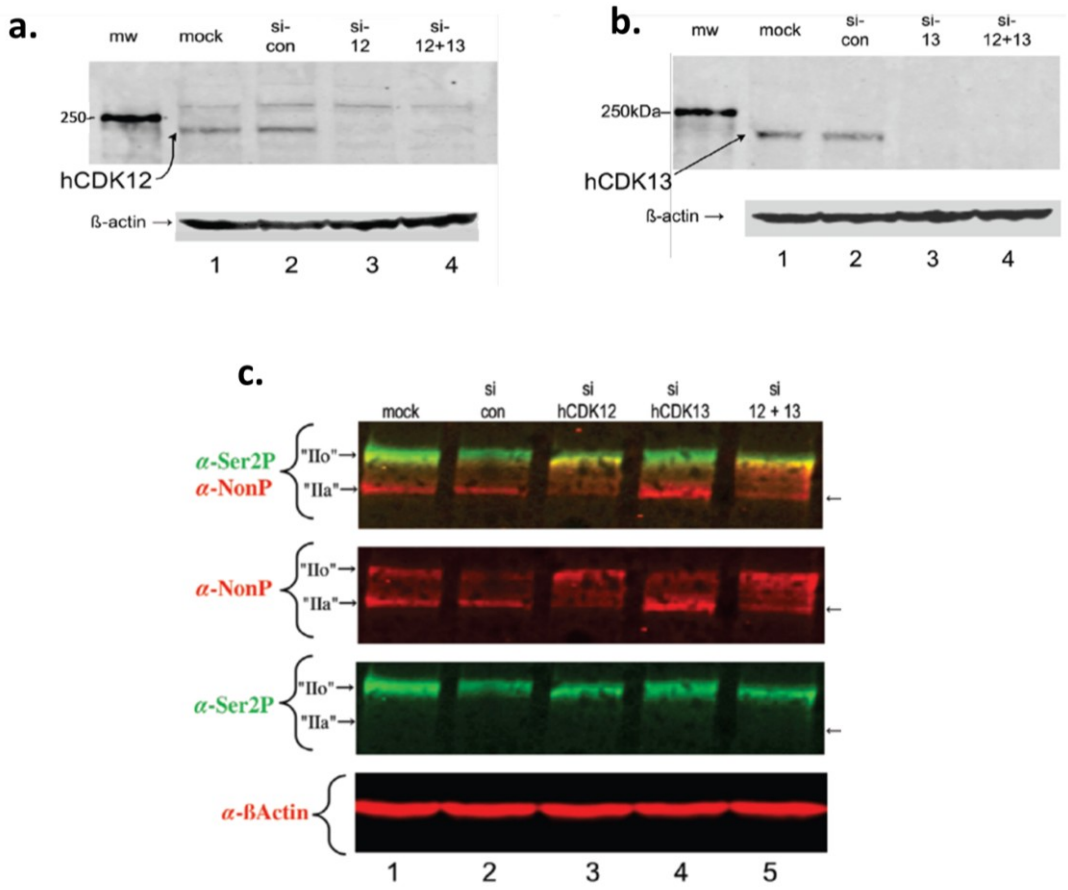

Figure 2.7: Western blots of HeLa whole cell extracts 72hr after RNAi knockdown of hCDK12 and/or hCDK13. Cells are treated with siRNA targeting hCDK12 (si-12), hCDK13 (si-13), or both (si-12+13). Mock and si-con are untreated and nontargeting controls; β-actin is used as a loading control. (a.) Anti-hCDK12 western confirming hCDK12 depletion (b.) Anti-hCDK13 western confirming hCDK13 depletion. (c.) Anti-Ser2P western using the 3E10 antibody (green) and anti-NonPhospho CTD repeat western using the 8WG16 antibody (red); note the changed reactivity of the 8WG16 antibody toward Rpb1 after hCDK12 depletion.

phosphorylated heptads (Chapman et al., 2007)) reveal that the Ser2 phosphorylated CTD signal undergoes a slight mobility shift (faster mobility) upon knockdown of hCDK12 (Fig. 2.7 c., In green). No discernable effect was seen after hCDK13 knockdown and an effect of similar magnitude to that of depleting hCDK12 alone is seen upon the depletion of both kinases. Fortunately, these unexpectedly middling results were

coupled with much more striking changes in the reactivity of the 8WG16 antibody, which specifically detects nonphosphorylated CTD heptad repeats (Jones et al., 2004) (Fig. 2.7 c., in red). Normally the 8WG16 antibody reacts with open epitopes in both the phosphorylated and unphosphorylated forms of the CTD (see Fig 2.7 c., α-NonP view only). However, hCDK12 depletion increased 8WG16 reactivity with the lower mobility IIo (hyperphosphorylated) band. One interpretation of these results is that hCDK12 depletion leads to an increase in unphosphorylated repeats (or Ser5 only phosphorylated repeats) in the elongating RNAPII. The decrease in phosphorylation is not large enough to drop the IIo band down to the IIa form, but is significant enough that the 8WG16 antibody reactivity towards the IIo band increases. Interestingly a similar 8WG16 detected Rpb1 mobility shift has been previously observed in Ctk1Δ yeast strains (Patturajan et al., 1999). Although subtle, the effect is highly reproducible, and has been also observed using HEK293T/17 cells. Therefore, the data from the knockdown experiments support the idea that hCDK12 is a bona fide CTD kinase in human cells. Because knockdown of hCDK13 does not appear to cause significant changes in global CTD phosphorylation, we cannot currently conclude that hCDK13 is a CTD kinase *in vivo*; however, *in vitro*, it clearly manifests CTD kinase activity.

2.4 Discussion

Through the co-transcriptional phosphorylation of its heptad repeats, the CTD of RNAPII coordinates transcription with a number of nuclear processes such as mRNA

processing and chromatin modification. Although *Drosophila melanogaster* CDK12 and human CDK12 and CDK13 have been previously identified by phylogenetic approaches as the closest homologs of the major transcription elongation-associated yeast Ser2 CTD kinase Ctk1 (Guo and Stiller, 2004; Liu and Kipreos, 2000), their status as functional CTD kinases was unclear. In this study we show that *Dm* CDK12 and human CDK12 are CTD kinases; the metazoan orthologs of yeast Ctk1.

In situ localization of dCDK12 on *Dm* polytene chromosomes presents a dramatic genomewide co-occurrence of dCDK12 with hyperphosphorylated RNAPII. The dCDK12 kinase appears to be closely associated with Pol II0, both at loci expressed during normal development, but also at heat-shock loci (Fig. 2.1). In contrast, the chromosomal distributions of dCDK12 and the Bur1 related Ser2 CTD kinase dP-TEFb are distinct (Fig. 2.2). In addition, ChIP studies show that whereas the amount of dP-TEFb peaks early in the gene (with paused polymerase), dCDK12 amounts peak later in the transcription unit (Fig. 2.3), resembling the behavior of yCtk1. Because the functions of P-TEFb, such as releasing RNAPII from elongation arrest early in transcription (Peterlin and Price 2006), are reflected in its localization patterns, we propose that the chromosomal localization patterns of dCDK12 similarly reflect its functions during transcription.

The fact that dCDK12 displays CTD kinase activity *in vitro* supports the idea that the primary function reflected in its chromosomal distribution is phosphorylation of the

CTD on elongating RNAPII. Given that hCDK12 also displays *in vitro* CTD kinase activity, CTD phosphorylation on elongating RNAPII is also the most likely function of the human homolog hCDK12. The changes in kinase distributions across transcription units, combined with ChIP data from Boehm et al. (Boehm et al., 2003), support a model in which P-TEFb prepares the polymerase for productive elongation near the promoter (through pause release, which occurs within 100bp of the TSS, and priming of the CTD for further phosphorylation (Qiu et al., 2009)) and then disassociates, allowing CDK12 to take over as the primary transcription elongation associated CTD kinase.

Knockdowns of dCDK12 and hCDK12 support the hypothesis that these kinases phosphorylate the CTD of Rpb1 *in vivo*. In *Drosophila* cells, which contain only one Ctk1 ortholog, the results are dramatic; over two-thirds of Ser2 phosphorylation, as detected by the mAb H5, is lost after knockdown of dCDK12 (Fig. 2.6 b.). Although these results are striking, it should be noted that despite being generally accepted by the *Drosophila* transcription field, the H5 antibody is raised against consensus sequence Ser2 phosphorylated CTD repeats. The *Drosophila* CTD only contains 2 consensus repeats and it is not known which, if any, of the phosphorylated nonconsensus repeats is detected by H5 (Fig 1.1). Thus, the precise epitope(s) of the H5 antibody in *Drosophila* are unknown. However, even with this caveat in mind, the results offer very strong evidence in favor of our initial hypothesis.

Looking at total RNAPII in whole-cell extracts of human cells, we also observe a change in CTD phosphorylation state after hCDK12 knockdown (Fig. 2.7 c.). As previously discussed (Munoz et al., 2010; Phatnani and Greenleaf, 2006) (also see 1.2.1), it is impossible to obtain a detailed understanding of phosphorylation patterns from antibody reactivities, but in this particular context, the important finding is that knockdown of hCDK12 results in alteration of the phosphorylation status of the CTD. Although the increase in the 8WG16 non-phospho antibody reactivity after knockdown of hCDK12 (Fig. 2.7 c.) may at first seem counterintuitive, it can be explained by the appearance of non-phospho epitopes within the repetitive CTD structure; an analogous observation was made previously in ctk1Δ strains of yeast (Patturajan et al., 1998a). Why the subtle, especially when compared to the dCDK12 knockdown results, change in CTD phosphorylation in human cells? Although we do not have a clear answer, it is worth repeating that the analysis of CTD phosphorylation states via western blotting of whole cell extracts is extremely complicated due to the presence of multiple forms of post-translationally modified CTDs in both the IIa and IIo bands (Munoz et al., 2010) and the limitations of the anti-phosphoCTD antibodies whose reactivity is affected by the modifications of neighboring residues and the residues in neighboring heptad repeats. Possible explanations include the previously mentioned divergence of the *Dm* CTD or the prospect of compensating kinase activities in human cells (e.g. P-TEFb may increase in promiscuity upon dCDK12 knockdown). Nevertheless, our results clearly

show that the CTD phosphorylation state is affected by hCDK12 knockdown which supports the hypothesis that hCDK12 is a functional orthologue of yCtk1 in humans. Tantalizingly, the knockdown experiments suggest that hCDK12 and hCDK13 play different functions *in vivo*, although hCDK13 does exhibit CTD kinase activity *in vitro*.

Overall we demonstrate that dCDK12 is a major elongation-associated CTD kinase in Drosophila, the ortholog of yCtk1. Furthermore, our results support the proposal that, in human cells, hCDK12 is a functional counterpart of yCtk1; hCDK13 is clearly a CTD kinase *in vitro*, but its functions *in vivo* are not yet clear. Our findings help clarify the relationships between two yeast CDKs, Ctk1 and Bur1, and their metazoan homologues and, most importantly, they draw attention to major metazoan CTD kinase activities that have gone unrecognized and unstudied until now. The identification of these elongation-phase Ser2 CTD kinases in metazoa enables us to study the link between CTD phosphorylation and transcription elongation-associated processes in higher eukaryotes. Therefore, further study of CDK12 and CDK13 promises to greatly expand our knowledge of both CTD phosphorylation and RNAPII transcription.

3. Identification of hCDK12 and dCDK12 Associated Proteins

3.1 Introduction

The majority of Ser2 CTD phosphorylation on productively elongating RNAPII in yeast (*Saccharomyces cerevisiae*) appears to be catalyzed by CTDK-I, a three subunit cyclin dependent kinase consisting of Ctk1 (a CDK homolog), Ctk2 (a cyclin homolog), and Ctk3 (function unknown). Ctk1 is not the only Ser2 CTD kinase in yeast, the essential Bur1 kinase (Bur1 is a CDK homolog and its cyclin partner is Bur2) has also been shown to contribute to Ser2 phosphorylation. Likewise, two primary Ser2 position CTD kinases have been identified in higher eukaryotes: The Bur1 related P-TEFb, composed of CDK9 and cyclinT, and the CTDK-I related CDK12/CyclinK complex. P-TEFb is recruited near the 5′ end of the transcription unit and in addition to Ser2 of the CTD it is able to phosphorylate Negative Elongation Factor (NELF) and the elongation factor Spt5, making it essential for entry into productive transcriptional elongation. The more recently discovered CDK12/CyclinK, on the other hand, appears to act further downstream within the transcription unit and currently its best studied function has been the phosphorylation of the Ser2 position of the CTD (Bartkowiak et al., 2010; Blazek et al., 2011).

Interestingly, human CDK12 (hCDK12) and *Dm* CDK12 are uncharacteristically large with molecular weights in excess of 160 kDa (most CDKs are about 50 kDa) and contain large N and C terminal extensions that flank the CDK kinase domain (Fig. 1.5).

These extensions contain multiple stretches of low complexity sequence and RS domains (domains characteristic of splicing factors believed to mediate protein-protein interactions (Zhong et al., 2009)); as expected of RS domain containing proteins, hCDK12 is found in nuclear speckles, which are believed to be storage sites for splicing factors (Zhong et al., 2009; Ko et al., 2001).

Much akin to their cell cycle regulating counterparts, the transcriptional CDK's activity depends on their interaction with their regulatory cyclin partners. Because members of the cyclin L class also colocalize to nuclear speckles and contain RS domains, it was postulated that they would be good candidates for the cyclin partners of hCDK12 and hCDK13. In order to test this proposal, previous studies overexpressed myc-tagged hCDK12 and hCDK13 in tandem with flag-tagged hCyclinL in HEK293 cells. When the overexpressed CDKs were immunoprecipitated, they were found to pull down the overexpressed hCyclinL (Chen et al., 2006; Chen et al., 2007).

Beyond the proposed interactions with CyclinL, no other CDK12 associated proteins have been identified. Due to the presence and sequence composition of CDK12's terminal extensions we hypothesized that CDK12 would mediate numerous protein-protein interactions which were likely to play a role both in the proper recruitment of CDK12 to the elongation complex and "non-catalytic", structure dependent, CDK12 functions. In order to gain more insight into the *in vivo* activity of hCDK12 and to inform future studies we set out to purify endogenous hCDK12 and

dCDK12 and to identify associated complexes and proteins using mass spectrometry. Functional annotation of the mass spectrometry identified CDK12 associated candidates reveals a strong enrichment of mRNA processing factors, nucleosome remodeling complexes, mRNA binding proteins, and nuclear speckle components. These results suggest that CDK12 is involved in multiple aspects of pre-mRNA processing, as is its yeast ortholog, Ctk1. Therefore we propose that CDK12 may affect RNA processing events in two distinct ways: indirectly through generating factor-binding phospho-epitopes on the CTD of elongating RNAPII, and directly through binding to specific factors.

3.2 Materials and methods

3.2.1 Antibodies and western blotting

rAP anti-hCDK12: Affinity-purified rabbit IgGs directed against a peptide comprising amino acids 201-220 of hCDK12; note that this antibody is distinct from the antibody used previously (see 2.2.1). rAP anti-dCDK12: Affinity-purified rabbit IgGs directed against a peptide comprising amino acids 705–722 of dCDK12 (CG7597; NP_649325.2). Anti-Ser2P: Mouse mAb (H5) from Covance. gAP anti-dRpb2: Affinity-purified goat IgG against a segment from dRpb2 (amino acids 519–992) (Skantar and Greenleaf, 1995). Western blot analysis was performed using the Odyssey infrared scanner and secondary antibodies from Li-Cor.

3.2.2 Identification of cyclin component of dCDK12

For dCDK12 cyclin identification, dCDK12 was purified as described previously (see 2.2.4.1). 15 μL of dCDK12 bead slurry was subjected to PAGE using a 4%–15% SDS gel (Biorad). After staining with Coomassie blue, gel bands of interest were excised and submitted for LC MS/MS analysis at the Duke Proteomics Facility (http://www.genome.duke.edu/cores/proteomics).

3.2.3 RNAi depletion of dCyclinK in S2 cell culture and qPCR

Primer design and RNAi treatment of S2 cell culture was performed as described previously (see 2.2.6). dsRNA construct primers were as follows:

CycKF 5′- GAATTAATACGACTCACTATAGGGAACAAATTCCTTCCAACGACG-3′

and

CycKR 5′- GAATTAATACGACTCACTATAGGGAACAGGTCCAGAACTTGGTGG-3′.

RNA was isolated using the RNeasy mini kit (Quiagen) (The optional on-column DNAse step was performed) and cDNA synthesis was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) using 1.5 μg of total RNA as input.

Quantitative PCR (qPCR) was performed using Power SYBR Green PCR master mix (Applied Biosystems) and the StepOnePlus real-time PCR system (Applied Biosystems); each sample was normalized based on the amount of Rp49 and changes in

gene expression were calculated using the ΔΔCt method. The forward and reverse primers used to detect each target are listed below.

CDK12: AGCTTCCGGTTCGCATCACTATCA & ATCTCTTGACGTTGGTGATCGGCT

CyclinK: AAACCTGGACCACACAAAGC & GCCCACATCAAAGATGAACC

Rp49: CCCAAGGGTATCGACAACAGA & CGATGTTGGGCATCAGATACTG

3.2.4 Immunopreciptation of dCDK12

2 L of *Drosophila* S2 cells were grown to $8x10^6$ in a spinner flask in M3 + BPYE with 10% serum. Nuclei were extracted following the method of (Weake et al., 2009), all steps were performed at 4 °C with protease inhibitors and 1 mM DTT present in all buffers. Briefly, cells were pelleted, subjected to a hypotonic swell, and disrupted with 40 strokes with a dounce homogenizer using the tight pestle; greater than 90% lysis was confirmed via microscopy. Nuclei were then pelleted, washed, and resuspended in a 350 mM NaCl, 0.1% Triton X-100 extraction buffer, and extracted for 1 hour (with end over end rocking). Nuclear extracts were cleared via centrifugation and stored at -80 °C until use; protein concentration was determined to be ~7 mg/mL by Bradford assay (BioRad).

The nuclear extracts were diluted to ~4 mg/mL total protein and the salt concentration was adjusted to 400 mM NaCl (all steps are performed at 4 °C with protease inhibitors (Sigma P8340) and 1 mM DTT present in all buffers). The dilute solution was cleared via centrifugation (20,000xg at 4 °C for 30 min) and 1 mL was used

for each pulldown. SoftLink Soft Release Avidin Resin (Promega), which was selected due to its low background binding, was saturated with a biotinylated goat anti-rabbit IgG antibody (Jackson ImmunoResearch Laboratories, Code #111-065-008). 10 µL of the resulting anti-rabbit IgG resin was used to preclear each 1 mL aliquot of the diluted S2 cell nuclear extract (1 hour of end over end rocking, 4 °C). The rest of the anti-rabbit IgG resin was either saturated with the rabbit anti-dCDK12 antibody or with total rabbit IgGs (purified using protein A agarose from rabbit pre-immune serum) as a negative control; 20 µg of antibody was used per 10 µL of the anti-rabbit IgG resin. The resulting bead-antibody-antibody complexes were then crosslinked using dimethyl pimelimidate (DMP) as described in Antibodies: A Laboratory Manual (Harlow and Lane, 1999). Following crosslinking, the beads were thoroughly washed with 1x PBS, 0.02% Tween20. The crosslinked anti-dCDK12 and total IgG beads were then used for overnight pulldowns using the precleared nuclear extracts as input (10 µL of the resin was used per 1 mL of nuclear extract).

Following the immunoprecipitation each 10 µL set of beads was washed 5x 1 mL of 400 mM NaCl extraction buffer, 1x 1 mL of 1x PBS with 1% NP-40, 3x 1 mL 1x PBS 0.02% Tween20 and 5x 1 mL 1x PBS. The beads were spun down, supernatant was completely removed, and the beads were eluted two times with 50 µL of 50 mM ammonium bicarbonate with 0.15% Rapigest detergent, for 5 min at room temperature with agitation per elution. The supernatant was then removed and pooled for each

sample (total rabbit IgG control and anti-dCDK12) and submitted to the Duke Proteomics Facility for subsequent analysis (LC MS/MS).

3.2.5 Immunoprecipitation of hCDK12

Dignam-Roeder nuclear extracts (Carey et al., 2009) (protein concentration at ~16 mg/mL) of HeLa cells were kindly provided by Dr. Irina Evsyukova from Dr. Mariano Garcia-Blanco's lab. The nuclear extracts were diluted to ~4 mg/mL total protein and the salt concentration was adjusted to 400 mM NaCl (All steps are performed at 4 °C with protease inhibitors (Sigma P8340) and 1 mM DTT present in all buffers). The dilute solution was cleared via centrifugation (20,000xg at 4 °C for 30 min) and 2 mL was used for each pulldown. For RNAse treatment, the 2 mL of nuclear extract was treated with 60 µL of RNAse cocktail (Ambion AM2246) or with 10 µL of RNAse inhibitor (RNAse inhibitor, Murine, NEB) and incubated at room temperature for 30 min before proceeding with the preclear and immunoprecipitation steps. RNAse amounts were determined in a pilot experiment using a spiked in 1000 bp RNA, which after RNAse treatment was phenol chloroform extracted and visualized via EtBr agarose gel electrophoresis to monitor RNA degradation.

SoftLink Soft Release Avidin Resin (Promega), which was selected due to its low background binding, was saturated with a biotinylated goat anti-rabbit IgG antibody (Jackson ImmunoResearch Laboratories, Code #111-065-008). 10 µL of the resulting anti-rabbit IgG resin was used to preclear each 2 mL aliquot of the diluted HeLa nuclear

extract (1 hour of end over end rocking at 4 °C). The rest of the anti-rabbit IgG resin was either saturated with the rabbit anti-hCDK12 antibody or with total rabbit IgGs (purified using protein A agarose from rabbit pre-immune serum) as a negative control; 20 µg of antibody was used per 10 µL of the anti-rabbit IgG resin. The resulting bead-antibody-antibody complexes were then crosslinked using dimethyl pimelimidate (DMP) as described in Antibodies: A Laboratory Manual (Harlow and Lane, 1999). Following crosslinking the beads were rinsed with 100 mM formic acid and thoroughly washed with 1x PBS, 0.02% Tween20. The crosslinked anti-CDK12 and total IgG beads were then used for overnight pulldowns using the precleared HeLa nuclear extracts as input (10 µL of the resin was used per 2mL of nuclear extract).

Following the immunoprecipitation each 10 µL set of beads is washed 5x 1 mL of IP buffer (Dignam-Roeder buffer D supplemented to 400 mM NaCl, see (Carey et al., 2009)), 1x 1 mL of 1x PBS with 1% NP40, 3x 1 mL 1x PBS 0.02% Tween20 and 5x 1mL 1x PBS. The beads were spun down, supernatant was completely removed, and the beads were eluted with 100 mM formic acid (100 µL of 100 mM formic acid per 10 µL of beads, 5 min, room temperature, with agitation). The supernatant was then removed and 200 µL of each sample (total rabbit IgG control and anti-hCDK12) was submitted to the Duke Proteomics Facility for lyophilization and subsequent analysis (LC MS/MS). For SDS-PAGE analysis of elution fractions the formic acid was neutralized using fresh 1 M NaOH.

3.3 Results

3.3.1 Identification of dCDK12's cyclin subunit[1]

Because *Drosophila* CDK12 (CG7597) had not been previously studied, we were interested in identifying its endogenous (not overexpressed) cyclin partner. dCDK12 was again immuno-affinity purified from Kc cell nuclear extracts as described in the preceding section (see 2.2.4.1); the resulting bead slurry was exhaustively washed, boiled in SDS loading buffer, and loaded onto an SDS-PAGE gel (Fig 3.1). Although we were wary of the methods used in the studies characterizing CyclinL as CDK12's partner (CDKs have been shown to be promiscuous for different cyclins (Fu et al., 1999) and tandem overexpression of both a cyclin and a CDK in the same system could "force" a non-native CDK/cyclin interaction) we still expected the cyclin subunit of dCDK12 to be dCyclinL. Since dCyclinL has a calculated MW of 63 kDa and RS domains, which tend to decrease mobility in SDS-PAGE gels due to their unique sequence compositions, bands "c," "d," and "e" were cut from the gel and submitted for protein identification by LC MS/MS mass spectroscopy (Duke Proteomics Facility) (Fig. 3.1). Gel slice "a" was also submitted for mass spectroscopy as it contained two distinct protein bands, as was gel slice "b," which contained strong protein bands and the dCDK12 protein as determined by western blotting (Fig. 3.1).

---

[1] The initial cyclin identification is adapted from, and published in, Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., Greenleaf, A. L., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24, 2303-16.

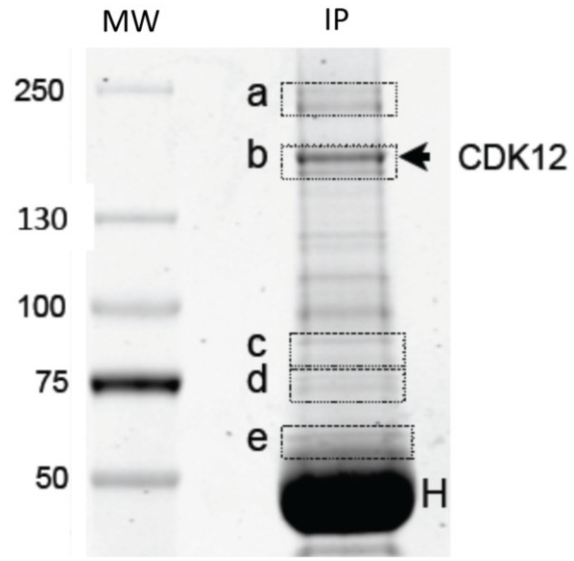

Figure 3.1: dCDK12 immunoprecipitation for cyclin partner identification. Coommasie-stained gel with molecular weight marker shown on the left (MW) and immunoprecipitation on the right (IP). H designates anti-dCDK12 antibody heavy chain. dCDK12 position (as determined by western blotting) is marked on right, and bands excised for mass spectroscopy are labeled a through e.

The mass spectroscopy results confirmed the presence of dCDK12 (28 tryptic peptides identified, 33% coverage of the protein, and ''100% probability'' of correct identification) and no other CDKs were identified in any of the excised bands. Surprisingly, the only cyclin identified in this analysis was *Drosophila* Cyclin K (dCyclinK) (Isoform a, in gel slice "c"; two tryptic peptides, ''100% probability'' of correct identification).

In order to further validate our result we used the NCBI non-redundant protein database to perform a BLAST search for metazoan sequences most similar to the yCtk2 protein, the cyclin subunit of yCtk1. The top ranked hits were all cyclins of the K class, implicating dCyclinK as a probable cyclin partner for dCDK12. We also searched the

DroID database of gene and protein interactions in *Drosophila*, for proteins that interact with dCDK12 (Yu et al., 2008). We were excited to discover that a DroID yeast two-hybrid screen picked up dCyclinK as the highest scoring interaction partner for dCDK12, out of 7 other (non-cyclin) positive hits (Finley Laboratory YTH results). Thus, our results suggest that dCDK12's cyclin partner is dCyclinK.

In order to confirm that dCyclinK was the actual cyclin partner of dCDK12 we used siRNA to deplete dCyclinK from *Drosophila* cell culture. Three dsRNA constructs were used: Anti CycK, anti E2 CDK12 (see 2.2.6 and Fig. 2.6) as a positive control, and anti lac-Z dsRNA obtained from Dr. Adelman's lab was used as a negative control (mock). The transfected cells were incubated for 72 hours and then harvested for western blotting of whole cell extracts. Although there is no antibody available for dCyclinK, we were able to confirm successful knockdown of both dCyclinK and dCDK12 at the mRNA level via qPCR (Fig. 3.2). In order to determine global CTD phosphorylation levels after dCyclinK depletion, we performed western blot analysis of dsRNA treated *Dm* whole cell extracts using the anti-Ser2P CTD antibody H5. Knockdown of dCyclinK resulted in an almost complete loss of H5 antibody reactivity, exactly analogous to dCDK12 knockdown. These results are clearly indicative of a CDK-cyclin partnership between dCDK12 and dCyclinK.

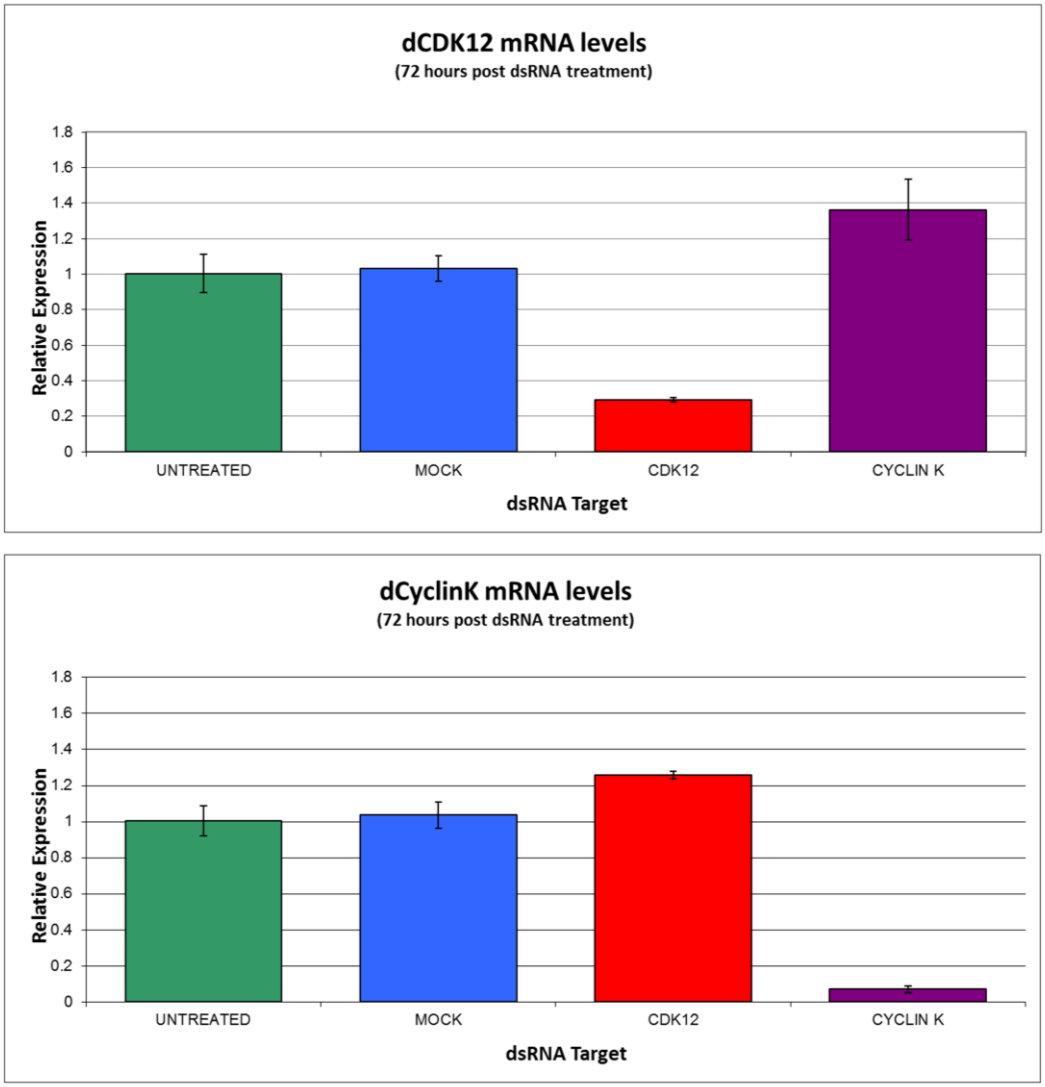

Figure 3.2: Quantification of dCDK12 and dCyclinK mRNA levels in *Drosophila* S2 cells 72 hours post dsRNA treatment using qPCR. Expression values are standardized to the untreated sample whose value was set at 1. Error bars are +/-2 standard errors of the mean, n=3.

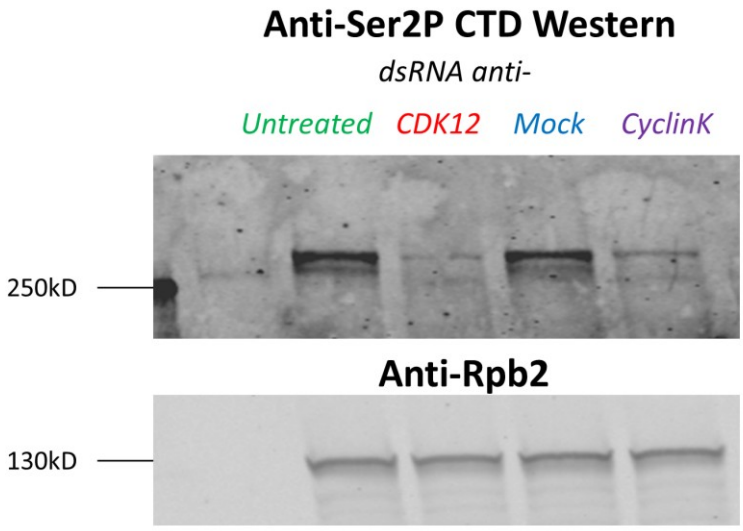

Figure 3.3: Western blots of Drosophila S2 whole cell extracts 72 hours post dsRNA treatment. Anti-Ser2P western using the H5 antibody (top); dRpb2 is used as a loading control(bottom). Mock cells are treated with anti-LacZ dsRNA.

3.3.2 Identification of dCDK12 associated proteins via mass spectrometry

In the course of our identification of dCyclinK as the cyclin partner of dCDK12 we also identified several other proteins of interest which co-precipitated with dCDK12 from *Drosophila* Kc cell nuclear extracts (Fig. 3.1), including components of the Brahma complex (dBRM); a SWI/SNF related chromatin remodeling complex previously shown to facilitate global transcription by, and to colocalize with, RNAPII (Armstrong et al., 2002). In order to confirm these results and to further explore dCDK12 interacting proteins, we immunopurified dCDK12 from *Drosophila melanogaster* S2 nuclear extracts using an affinity purified anti-dCDK12 antibody tethered to streptavidin beads. Immunopurified complexes were eluted and analyzed by mass spectrometry (LC MS/MS; Duke Proteomics Facility). Removing proteins present in control

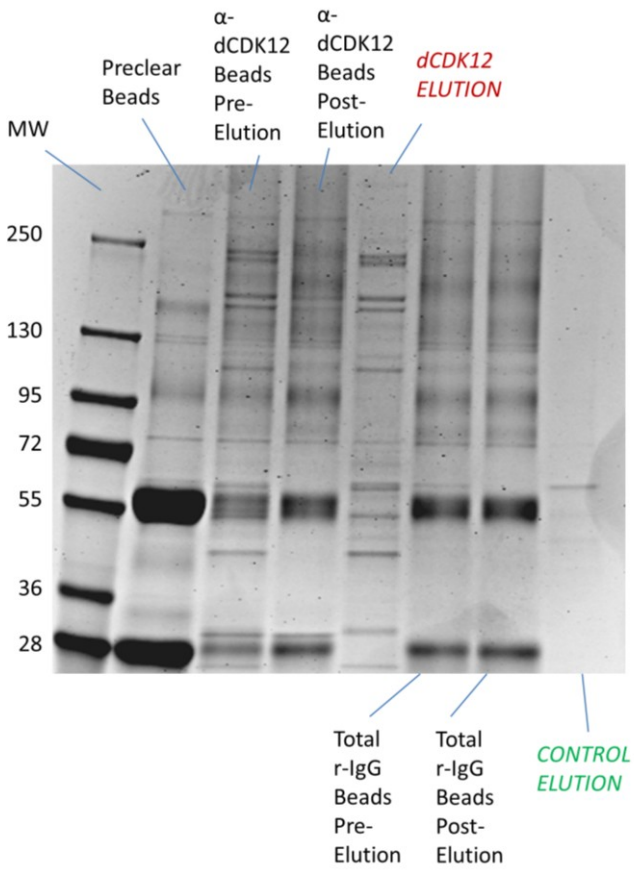

Figure 3.4: Coomassie-stained gel of dCDK12 immunoprecipitations and elutions submitted for mass spectrometric analysis. The "Preclear Beads" lane contains uncrosslinked, biotinylated goat anti-rabbit IgG antibody, saturated beads used to preclear the nuclear extract. The "Pre-elution" and "Post-Elution" lanes contain crosslinked anti-dCDK12 or anti-total rabbit IgG antibody saturated beads, before and after elution. The "ELUTION" lanes are representative of the samples sent for analysis; dCDK12 is the top band of the doublet located at ~160kDa.

immunoprecipitations (using total rabbit IgG instead of anti-dCDK12 IgG) results in a final hit list composed of 81 proteins (See Appendix A) (Fig 3.4).

Functional annotation of this list using the DAVID database (Huang da et al., 2009a; Huang da et al., 2009b) reveals a strong enrichment for mRNA processing and chromatin remodeling factors. As expected we find almost every component of the dBRM complex coprecipitating with dCDK12; this includes Brahma, Moira, OSA,

Polybromo, and the various Brahma associated proteins or BAPs (BAP 60kD, BAP 170kD, BAP 11kD, and BAP 55kD). In the RNA processing category the most highly enriched DAVID annotation cluster was that of the exosome components; Rrp4, 6, 40, 45, and 46, Mtr3, and Ski6 were all present in the dCDK12 elutions. The exosome is a complex of exoribonucleolytic and RNA binding proteins involved in 3′ end RNA degradation and processing; much like dCDK12 and the dBRM complex, it is also recruited to actively transcribed loci and colocalizes with elongation factors. Additionally, the mass spectrometry identified eIF4A3, a component of the exon junction complex, and a number of proteins implicated in splicing, including the spliceosomal component Prp19. Intriguingly, the nuclear speckle localized SC35 splicing factor (also known as Serine/arginine-rich splicing factor 2 or SRSF2 in humans) was also identified, but was discarded from the final analysis as it was also detected in the control (albeit at lower levels; 3 peptides in dCDK12 versus 1 peptide in the total rabbit IgG pulldown).

3.3.3 Identification of hCDK12 associated proteins via mass spectrometry

In order to identify hCDK12 interacting proteins we also immunopurified CDK12 from Dignam Roeder nuclear extracts of HeLa cells (Carey et al., 2009) using an affinity purified anti-hCDK12 antibody tethered to streptavidin beads. Immunopurified complexes were eluted and analyzed by mass spectrometry (LC MS/MS; Duke Proteomics Facility). Filtering out proteins present in control immunoprecipitations

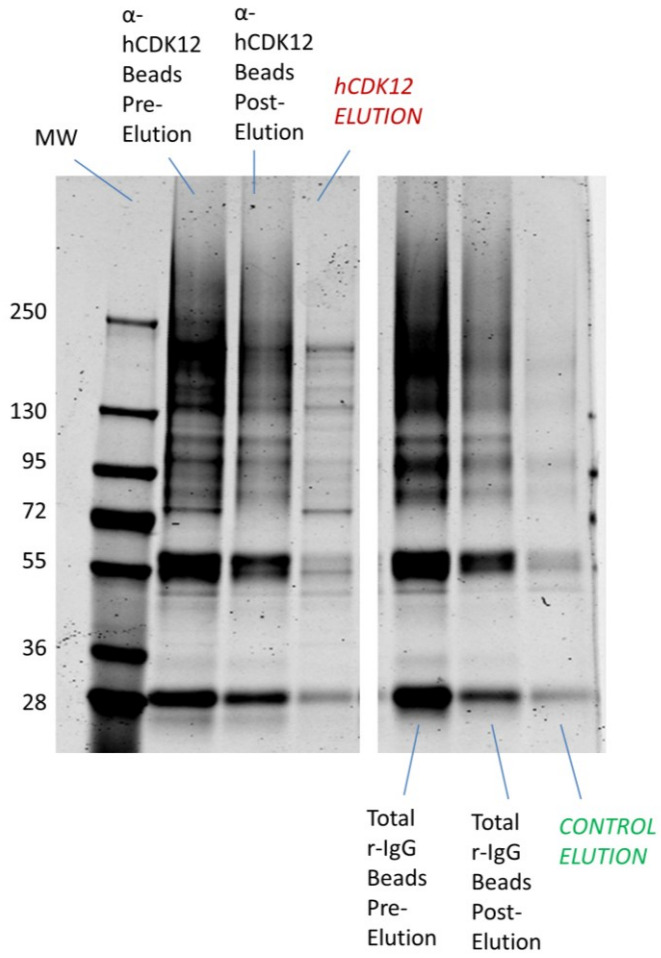

Figure 3.5: Coomassie-stained gel of hCDK12 immunoprecipitations and elutions submitted for mass spectrometric analysis. The "Pre-elution" and "Post-Elution" lanes contain crosslinked anti-hCDK12 or anti-total rabbit IgG antibody saturated beads, before and after elution. The "ELUTION" lanes are representative of the samples sent for analysis; hCDK12 is the prominent band located at ~200kDa in the "ELUTION" lane.

(again, using total rabbit IgG instead of anti-CDK12 IgG) results in a total of 309 identified proteins in the final hit list (See Appendix B) (Fig 3.5).

Functional annotation of the positive hits using the DAVID database (Huang da et al., 2009a; Huang da et al., 2009b) reveals a strong enrichment of mRNA splicing factors, mRNA binding proteins, and nuclear speckle components. In agreement with

this functional annotation, the presence of RS domains in hCDK12's N-terminus, and hCDK12's localization to nuclear speckles (Blazek et al., 2011; Ko et al., 2001) we find Serine/arginine repetitive matrix protein 1 and 2, and many SR Splicing Factors (SRSF1, 3, 4, 5, 6, 7, 8, 9, and 10) coprecipitating with hCDK12. In addition, the mass spectrometry identified several other proteins of interest including PCF11 (a 3' end processing factor), Prp19, several exosome components, and components of the exon-junction complex (eIF4A3, MGN, RNPS1, and RBM8; previously shown to interact with SR proteins (Singh et al., 2012)). In order to confirm these interactions and to determine whether they are mediated via RNA, the experiment was repeated again, this time incorporating an RNAse treatment step. Unfortunately the final elutions in this repeat experiment did not contain as much material (both in the RNAse treated and untreated samples) resulting in a much lower depth of coverage in the mass spectrometric analysis. Despite this we were still able to verify many of the above interactions and determine that they were not dependent on RNA: Of the proteins listed above, only Serine/arginine repetitive matrix protein 1 was found to be RNA dependent; PCF11, all of the components of the exon junction complex except for RBM8, and SRSF4, 6, 7, 9, and 10 repeated. Although this initial list of CDK12 associated proteins will require further validation, it affirms an interaction between CDK12 and other RS domain containing proteins and reinforces the role of CDK12 as an important actor in cotranscriptional mRNA processing.

3.4 Discussion

Intrigued by the N and C terminal arms which extend away from CDK12's kinase homology domain, we hypothesized that CDK12 was likely to be associated with a large number of accessory factors. Because these factors could potentially play a role in the targeting of CDK12 to the elongation complex and were likely to offer insights into potential non-catalytic roles of CDK12, we have identified proteins that co-immunoprecipitate with CDK12 from HeLa and *Dm* cell nuclear extracts under fairly stringent conditions.

Our experiments began with an effort to identify the cyclin partner of dCDK12, which was previously hypothesized to be CyclinL. Surprisingly mass spectrometric analysis of slices from an SDS-PAGE gel of immunopurified dCDK12 implicated dCyclinK as its potential partner. Followup analysis using dsRNA confirmed this interaction and several subsequent studies in human cells have determined that the cyclin partner of hCDK12 is also hCyclinK (Blazek et al., 2011; Cheng et al., 2012). Pertinently, human CyclinK was previously found to associate with CDK9 via a yeast two hybrid screen and was shown to stimulate CDK9's CTD kinase activity *in vitro* (Fu et al., 1999). Further investigation into the proposed CDK9/CyclinK complex implicated CyclinK as playing a role in replication stress and genome integrity, however these functions were attributed to CDK9 (Yu and Cortez, 2011; Yu et al., 2010). As several studies have now confirmed that the endogenous human CyclinK is actually non-

promiscuous and associates exclusively with CDK12 and CDK13 (Blazek et al., 2011; Cheng et al., 2012), the genome integrity functions of CyclinK are currently being reassessed from a CDK12 centered perspective (Bartkowiak et al., 2011; Blazek, 2012; Blazek et al., 2011; Joshi et al., 2014; Winsor et al., 2013). Exactly what role hCDK12/CyclinK plays in the response to DNA damage is still being investigated, but it promises to be a field of intense research in the coming years (see chapter 6 for more discussion).

In view of the SR protein-like "arms" of the CDK12 protein, it is perhaps not surprising that RNA processing factors are by far the most abundant functional class of CDK12-associating proteins. In the human pulldown, over 30 splicing proteins are represented, including SRRM1 & 2, TRA2A & B, U2AF2, and SRSF1, 3, 4, 5, 6, 7, 8, 9 & 10. Also present are components of the PRPF19-CDC5L complex, a number of DDX and DHX helicases, and members of the exon junction complex. Beyond splicing, several 3' end-formation factors are also represented including Pcf11 and numerous subunits of the RNA exosome, which are also heavily enriched in the *Drosophila* CDK12 pulldown. In terms of nucleosome remodeling we find that dCDK12 appears to have a strong interaction with the SWI/SNF related dBRM complex; although we do not observe this strong enrichment with hCDK12, the SWI/SNF complex subunit SMARCC1 and the SWI/SNF related SMCA5 are present in the human hit list.

On a final note of interest, mass spectrometry of the human pulldowns identified PAP1L and PABP4; putative human homologues (by Uniprot Blast) of the yeast Ctk1 interacting protein GBP2 (Ho et al., 2002). Yeast contains only three proteins with canonical SR structures (RNA recognition motifs and an RS dipeptide rich region): Gbp2, its paralog Hrb1, and Npl3. Intriguingly, all three of these proteins have been shown to interact with Ctk1 (Ho et al., 2002), suggesting an early evolutionary link between the Ser2 CTD kinases and the SR proteins.

Although the mass spectrometry results will require additional experimental validation, they suggest that CDK12 is involved in multiple aspects of pre-mRNA processing. As an example, in terms of splicing, we suggest that a number of SR proteins are recruited to the transcription elongation complex by virtue of RS domain interactions with CDK12. Therefore, we propose that CDK12 affects RNA processing events in two distinct ways: Indirectly through generating factor-binding phospho-epitopes on the CTD of elongating RNAPII, and directly through binding to specific factors via its extended N and C terminal domains.

4. Expression, Purification, and Characterization of the hCDK12/CyclinK Complex and Mutants

4.1 Introduction

The C-terminal domain (CTD) of RNA polymerase II undergoes a cycle of phosphorylation which allows it to temporally couple transcription with a wide array of transcription associated processes. Most of the Ser2 CTD phosphorylation that occurs during transcriptional elongation in the yeast *Saccharomyces cerevisiae* has been shown to be mediated by CTDK-I, an enzyme consisting of three subunits: Ctk1 (a CDK homolog), Ctk2 (a cyclin homolog), and Ctk3 (an essential accessory factor of unknown function). Ctk1 coexists with the essential Bur1/Bur2 complex (Bur1 and 2 are, respectively, CDK and cyclin homologs) which has also been shown to contribute to Ser2 phosphorylation and to prime the CTD for further phosphorylation near the 5′ end of the gene.

Correspondingly, two Ser2 position CTD kinases have also been identified in metazoa: The Bur1 related P-TEFb (composed of CDK9 and cyclinT) and the CTDK-I related CDK12/CyclinK complex. Like Bur1, P-TEFb is recruited near the 5′ end of the transcription unit, and in addition to phosphorylating Ser2 of the CTD, is involved in mediating entry into productive transcriptional elongation by releasing RNAPII from promoter proximal pausing. In contrast, CDK12/CyclinK acts towards the 3′ end of the transcription unit and currently its best studied phosphorylation target is the Ser2 position of the CTD (Bartkowiak et al., 2010; Blazek et al., 2011).

Recent studies have elucidated the structure of a truncated form of the CDK12/CyclinK complex and implicated the complex in the 3′ end processing of the MYC gene (Bosken et al., 2014; Davidson et al., 2014). In addition, hCDK12 has been identified as a tumor suppressor for ovarian cancer (Joshi et al., 2014); however, much is still unknown about the *in vivo* roles of CDK12/CyclinK. Given our analysis of dCDK12 and hCDK12 associated proteins, it is likely that CDK12/CyclinK plays both catalytic and noncatalytic (structural) roles in mRNA processing and other transcription related functions. Additionally, due to the fact that the sensitivity profile of hCDK12/CyclinK to CDK inhibitors has not been previously investigated and that P-TEFb inhibition/depletion has drastic consequences on global transcription (due to the inhibition of pause release near the TSS), it is challenging to determine which of the functions attributed to P-TEFb are actually due to the lack of downstream CDK12 activity. In an effort to further characterize CDK12s roles *in vivo,* and to help dissect its catalytic and non-catalytic functions, we have purified to near homogeneity full-length, active, human CDK12/CyclinK, created an analog sensitive CDK12 mutant, and characterized CDK12's substrates specificity and sensitivity to the "CDK9 specific" inhibitor flavopiridol (FVP).

4.2 Materials and methods

4.2.1 Antibodies and western blotting

rAP anti-hCDK12: Affinity-purified IgGs directed against a peptide comprising amino acids 201-220 of hCDK12 (NCBI RefSeq: NP_057591.2). The anti-CyclinK antibody was purchased from Abcam (ab130475). The anti-gp64 antibody against baculovirus envelope protein was purchased from eBioscience. Western blot analysis was performed using the Odyssey infrared scanner and secondary antibodies from Li-Cor.

4.2.2 GST-CTD substrates

GSTy-CTD was purified as previously described (Morris et al., 1997). Mutant GST-CTD substrate constructs (WT, S2A, S2E, S5A, and S5E (originally designed by Dr. Jeff Corden's lab (Patturajan et al., 1998a)) were kindly provided by Dr. Aseem Ansari's lab. The S7E construct was created by gene synthesis of DNA containing 16 CTD heptad repeats with each S7 position replaced with E, followed by an HA tag (YPYDVPDYA), a single glutamate and a stop codon (in order to be identical to the WT GST-CTD substrate originally designed by the Corden lab). The sequence also incorporates a BamHI and SalI site at the 5′ and 3′ end respectively (IDT). The S7E CTD sequence was cloned into the pGEX-5X-1 vector (GE Healthcare: Life Sciences) using the BamHI and SalI restriction sites. Expression of all constructs was performed in BL21-Codon Plus (DE3)-RIPL cells (expression conditions were 25 °C with 1 mM IPTG; Stratagene). After 16

hours of expression, cells were sonicated and the fusion protein was purified using a glutathione-agarose (Pierce) column following the manufacturer's protocol. Peak elution fractions were then pooled and submitted to size exclusion chromatography (Superdex 200 HR 10/30, Amersham). Peak size exclusion fractions were pooled and buffer exchanged (Amicon Ultra Centrifugal Filters) into 25 mM HEPES pH 7.6 with 50 mM NaCl to a final protein concentration of 0.5 μg/μL.

4.2.3 CTD kinase assays

CTD kinase assays were performed essentially as described previously (Jones et al., 2004), with minor variations: Kinase reactions included either 30 or 300 μM ATP, 10 μCi of [γ-^{32}P]ATP, 5 mM $MgCl_2$, 25 mM Tris pH 7.6, 5% Glycerol, 2 mM DTT, 150 mM NaCl, 1 μg of GST-yCTD, and 50-100 ng of hCDK12/CyclinK in a 25 μL reaction volume. Reactions were performed for 15-45 min (specified in each figure) at 37 °C. Reaction times and enzyme amounts were selected to be in the range where phosphate incorporation is linear with time. Phosphorylated products were visualized using film (Kodak Biomax MR) or a PhosphorImager; quantification was performed using a PhosphorImager and ImageQuant software. Ess1 was provided by Dr. Pei Zhou's laboratory. 1-NM-PP1 was purchased from Axon Medchem and dissolved in DMSO. For DRB and flavopiridol (FVP) inhibition, equal activity of CDK9, provided by the laboratory of Dr. David Price, and CDK12/CyclinK was used (0.6 μM final concentration of CDK9/CyclinT in the reaction mixture). FVP was purchased from Sigma Aldrich

(F3055) and DRB was provided by Dr. Shudong Wang from the University of Nottingham. Inhibitors were dissolved in DMSO and then aliquoted at the appropriate concentrations in 4% DMSO. 1 μL of either 4% DMSO or the appropriate concentration of each inhibitor in 4% DMSO was added to the reaction mixture for a final volume of 25 μL. Analysis of the inhibition data was performed using GraphPad Prism 6.

4.2.4 Construction of CDK12/CyclinK expression constructs and mutants

The cDNA for hCDK12 was purchased from Open Biosystems (Clone ID: 9021722; NCBI accession BC140854), as was the cDNA for the short, 40 kDa, form of hCyclinK (Clone ID: 3907416 NCBI accession BC015935). The full length hCyclinK cDNA was a very kind gift from Dr. Grace Cheng and Dr. Gregg Morin (Cheng et al., 2012).

The hCDK12 cDNA was then extended using PCR: The forward primer contained an upstream NarI extension (GTAGCAGGCGCCATGCCCAATTCAGAGAGACA) and the reverse primer annealed just upstream of the hCDK12 stop codon, replacing it with an extension containing a part of the P2A sequence with a BamHI restriction site (TTGCTTTAACAGAGAGAAGTTCGTGGCTCCGGATCCGTAAGGAACTCCTCTCCCTCTTC).

The hCyclinK cDNAs were also extended using PCR: The forward primer contained part of the P2A sequence

(CAAGCAGGAGACGTGGAAGAAAACCCCGGTCCTATGAAGGAGAATAAAGAAA
ATTCAAGC) and the reverse primer contained an extension containing a NotI restriction site (GTAGCAGCGGCCGCTTATCTCATCCAGGCTGCCC). This extended hCyclinK PCR product was then extended again using the reverse primer from above and another forward primer which added more of the P2A sequence, including the aforementioned BamHI restriction site:
GGATCCGGAGCCACGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGA.

The hCDK12 PCR product was then cloned into the pFastBac HT B (Invitrogen) vector using NarI and BamHI restriction sites. The resulting plasmid was further modified by the addition of the full-length or 40 kDa, fully extended, hCyclinK PCR products using the BamHI and NotI restriction sites; the resulting plasmids were the final pFastBac HT B hCDK12/CyclinK and the pFastBac HT B hCDK12/CyclinK(40kDa) expression constructs.

Mutations (D877N, K756A, and F813G) were introduced into the pFastBac HT B hCDK12/CyclinK expression construct using the QuikChange Site-Directed Mutagenesis Kit (Stratagene).

4.2.5 Expression and purification of hCDK12

Expression was performed essentially as described in the Bac-to-Bac Baculovirus Expression System instruction manual (Invitrogen). Miniprepped pFastBac HT B hCDK12/CyclinK plasmids were transformed into DH10Bac E. coli for transposition into

the expression bacmid. Transformed colonies were screened for the CDK12/CyclinK insert by PCR using M13 primers and a primer internal to hCDK12 (GACTGACCGACTGCCTTCTC). Bacmids were prepared from 150 mL DH10Bac cultures with the NucleoBond Xtra Midi kit (Macherey-Nagel) and transfected into *Spodoptera frugiperda* (Sf9) insect cells using CellfectinII (Life Technologies) following the manufacturer's protocol. Sf9 cells were maintained in Hyclone SFX-Insect Media (Thermo Scientific) at 27 °C.

Virus amplification was performed as described in the Bac-to-Bac manual for 3 rounds of amplification. Viral titer was determined via western blotting of the cell culture supernatant with the anti-gp64 antibody against baculovirus envelope protein (eBioscience). Expression was optimized by varying the amount of virus (using western blotting for gp64 for relative quantification) and time of expression; expression was monitored through western blotting for hCDK12. For purification, 500 mL of Sf9 cells at $1x10^6$ cells/mL were infected with baculovirus, incubated for 96 hours, and cells were collected by centrifugation at 730xg for 15 min at 4 °C (all subsequent steps at 4 °C), washed with ice cold 1x PBS, and pelleted again.

Cells were resuspended in cytoplasmic lysis buffer (10 mM HEPES, pH 8.0, 320 mM sucrose, 3 mM calcium chloride, 2 mM magnesium acetate, 1 mM DTT, 5 mM NaF, and 0.5% NP40) with protease inhibitor mixture for tissue extracts (P8340, Sigma). The resulting slurry was homogenized using 20 strokes with a tight fitting pestle of a dounce

homogenizer. Nuclei were then pelleted by centrifugation at 1,500xg for 15 min. The supernatant was removed and the nuclei were resuspended in 6 mL of nuclear resuspension buffer (20 mM HEPES, pH 8.0, 1.5 mM $MgCl_2$, 20 mM KCl, 25% glycerol, 1 mM DTT, 5 mM NaF, and protease inhibitor mixture for tissue extracts). Nuclei were then lysed by the dropwise addition (with stirring) of 6 mL of nuclear extraction buffer (80 mM HEPES, pH 7.6, 1.2 M NaCl, 20% glycerol, 0.06% TritonX-100, 1 mM DTT, 5 mM NaF, and protease inhibitor mixture for tissue extracts). 6 mL of PEG solution (18% PEG 8000, 1 M NaCl, 1 mM DTT, 5 mM NaF, and protease inhibitor mixture) was then added, and the solution was incubated at 4 °C for 45 min with rotation. The insoluble fraction was removed by centrifugation at 20,000x*g* for 30 min. The soluble lysate containing hCDK12/CyclinK was then adjusted to 20 mM Imidazole and bound in bulk to a preequilbrated Ni Sepharose 6 Fast Flow column (GE Healthcare) for 2.5 hours at 4 °C rocking end over end. The resin was collected and washed with 10 column volumes of high salt wash buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 20 mM imidazole). The protein was eluted with 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 250 mM imidazole, 1 mM DTT, 5 mM NaF, 8% glycerol, and protease inhibitor mixture. The elution peak from the Ni-Sepharose was pooled and, using a 500 µL loading loop, applied to an ÄKTA FPLC system equipped with a Superdex 200 10/30 HR column (GE Healthcare) at a flow rate of 0.25 ml/min with 25 mM HEPES pH 7.6, 8% Glycerol, and 300 mM NaCl as the running buffer. 500 µL fractions were collected and fractions containing

hCDK12/CyclinK were identified by SDS-PAGE and flash frozen for storage at -80 °C until use.

4.2.6 Phosphatase treatment

20 µL of a CDK12/CyclinK Ni-Sepharose elution fraction was treated with 10 units of FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) at 37 °C for 1 hour. The reaction was stopped by the addition of 5x SDS loading buffer (Laemmli buffer) and analyzed by SDS-PAGE.

4.3 Results

4.3.1 Construction of the hCDK12 expression construct

It has been previously reported in the literature that full length human CDK12 expressed in baculovirus-infected cells is highly insoluble or inactive (Bosken et al., 2014; Ko et al., 2001). We hypothesized that this may be due to the lack of appropriate levels of hCDK12's cyclin partner, which would be likely to affect the stability of the kinase. To address this problem we employed a "2A" peptide-linked multi-cistronic construct (Szymczak-Workman et al., 2012). The "2A" peptide sequence results in a site specific, sterically induced, cleavage of the fusion protein as it is produced on the ribosome: This allows for the expression of both hCDK12 and hCyclinK, in an exact one to one ratio, from a single open reading frame. The final construct contained the His-tagged hCDK12 transcript variant 2 sequence (Genbank accession NM_015083) lacking the final stop

codon (1481AA; transcript variant 1 [Genbank accession NM_016507] contains an extra 9AAs from exon 14, which map to the C-terminal arm of the kinase), the 2A sequence

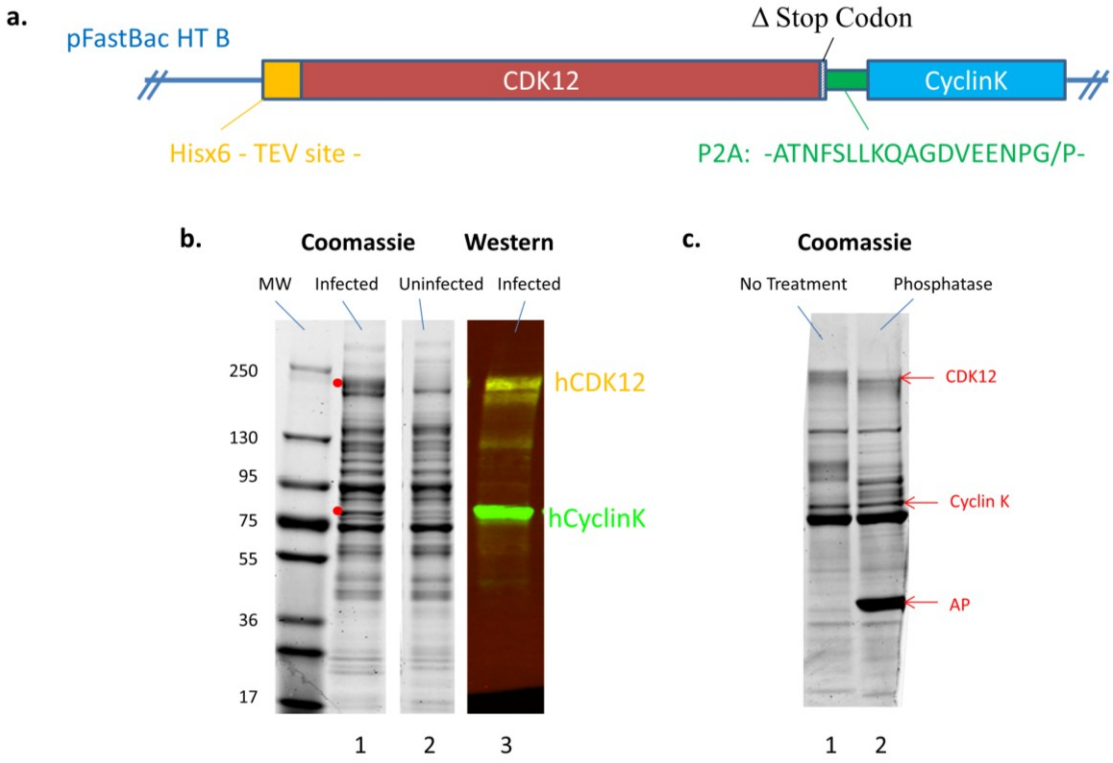

Figure 4.1: The hCDK12/CyclinK baculovirus construct and its expression. (a.) Schematic of the hCDK12/CyclinK expression construct, which consists of the full length human CDK12 gene (lacking its stop codon) fused to the P2A sequence from porcine teschovirus and followed by the full length sequence for hCyclinK. The complete construct was cloned into the MCS of the pFastBac HT B insect cell expression vector (Invitrogen), which led to the inclusion of a 6xHis tag followed by a TEV protease site at the N-terminal end of hCDK12. (b.) Coomassie stained gels and western blots of Ni column elution fractions of hCDK12/CyclinK baculovirus infected and uninfected Sf9 cell culture lysates. Infection of Sf9 cells with hCDK12/CyclinK baculovirus results in the appearance of two additional bands on the coomassie stained elution fractions (red dots). Western blot analysis of infected elution fractions with anti-hCDK12 (in yellow) and anti-hCyclinK (in green) antibodies indicate the presence of soluble hCDK12/CyclinK complex in the infected lysates. (c.) Coomassie stained gel of Ni column fractions of hCDK12/CyclinK baculovirus infected Sf9 cell lysates untreated and treated with Alkaline Phosphatase (AP).

from porcine teschovirus-1 (19AA), and the full sequence for CyclinK (580AA) (Fig. 4.1 a.) (Cheng et al., 2012). Co-expression of both the CDK and its cyclin partner in baculovirus-infected Sf9 cells followed by Ni Column purification resulted in the appearance of two new bands in coomassie gels of the elution fractions; a diffuse band at ~180 kDa and a sharp band at ~75 kDa. These bands were confirmed to be CDK12 and CyclinK by western blotting (Fig. 4.1 b.). The diffuse nature of the CDK12 band is likely due to a high degree of post translational modification; phosphorylation of CDK12 has been reported previously (Ko et al., 2001) and phosphatase treatment of Ni column elutions of CDK12 resulted in a mobility shift and partial collapse of the CDK12 coomassie band (Fig. 4.1 c.). Therefore, co-expression of CDK12 with its cyclin partner alleviates solubility issues and is a viable strategy for the production of recombinant enzyme.

4.3.2 Purification and kinase assays of WT hCDK12, kinase dead, and analog sensitive mutants

His-tagged CDK12 was purified from baculovirus infected cells via isolation of nuclei, followed by nickel column chromatography and size exclusion chromatography (see 4.2.5). Serendipitously, the large size of the CDK12/cyclinK complex, coupled with the high likelihood of non-globular structural elements in the N and C-terminal arms of its components, led it to elute near the void volume of the size exclusion column. Therefore, a simple two-step purification strategy yielded highly purified complex

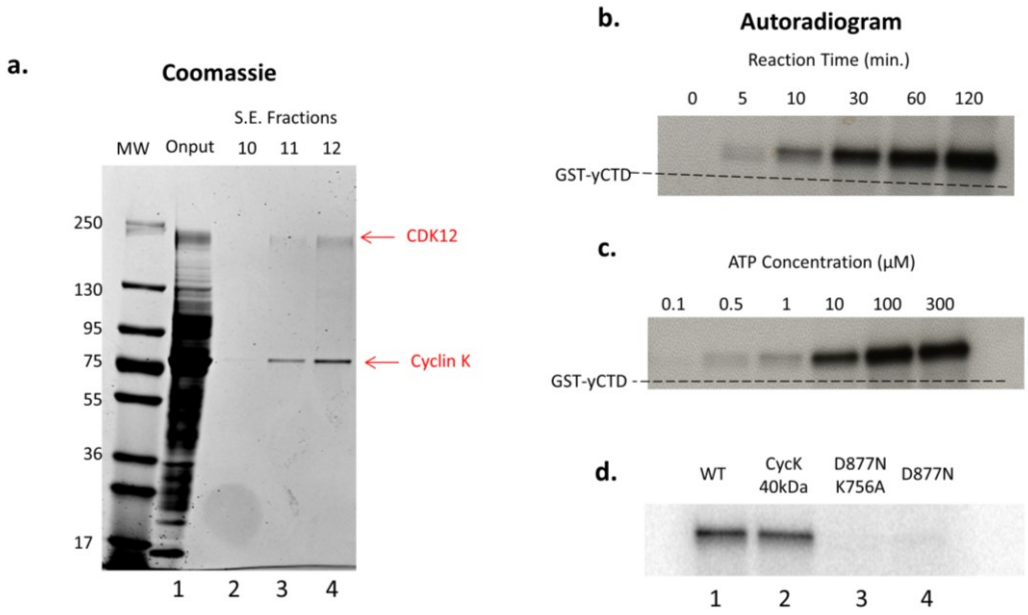

Figure 4.2: Purification and CTD Kinase activity of hCDK12/CyclinK complex and mutants. (a.) Coomassie stained gel of size exclusion chromatography (Superdex 200 HR 10/30) fractions from a hCDK12/CyclinK purification. Onput (lane one) consists of pooled Ni column elution fractions of hCDK12/CyclinK baculovirus infected Sf9 cell nuclear extracts. The majority of fraction #10 (lane two) consists of column void volume while fraction #11 (lane three) is the first true elution fraction. (b.) Purified hCDK12/CyclinK CTD kinase activity was assayed in the presence of [γ-^{32}P]ATP using 1 ug of a GST-yCTD fusion protein (26 heptad repeats) as substrate and reaction mixtures were analyzed by SDS-PAGE. Autoradiograms of the SDS-PAGE gels revealed a time (ATP at 300 μM) and (c.) ATP dependent (reaction time at 45 min) CTD kinase activity. The SDS-PAGE mobility of the unphosphorylated GST-yCTD substrate is indicated via a dashed line. (d.) Autoradiograms of CTD kinase assays using purified mutants of the hCDK12/CyclinK complex. Equal amounts of each construct are used in the reactions; hCDK12/CyclinK complex containing a shortened form of CyclinK exhibited comparable activity to full length wild-type complex (compare lanes one and two), while a kinase dead double (lane 3) and single mutant (lane 4) exhibited little to no residual activity.

which could be employed in subsequent studies (Fig. 4.2 a.). Purified CDK12/CyclinK exhibited a time and ATP dependent CTD kinase activity when assayed using GST-yeastCTD (GST-yCTD) fusion proteins as substrate in the presence of [γ-^{32}P]ATP (Fig. 4.2 b. and c.). Unlike CTDK-I, which is inhibited by the presence of additional salt in the kinase buffer, CDK12/CyclinK activity was stimulated by salt (NaCl and KCl were tested at 0, 150, and 200 mM) and strongly inhibited by detergents (0.05% TritonX-100, NP40,

and Tween20 were tested) (Lee and Greenleaf, 1989) (Fig. 4.3). We also tested hCDK12/CyclinK activity towards the GST-yCTD in the presence of the yeast proline isomerase Ess1, but observed no stimulation of activity (Fig 4.4). In addition to the full length construct, we also constructed and purified a CDK12/CyclinK complex containing the originally identified 40 kDa form of CyclinK lacking a C-terminal, proline rich domain (this isoform was subsequently proven to be an artifact) (Bartkowiak et al., 2010; Blazek et al., 2011; Cheng et al., 2012; Edwards et al., 1998); however we could not

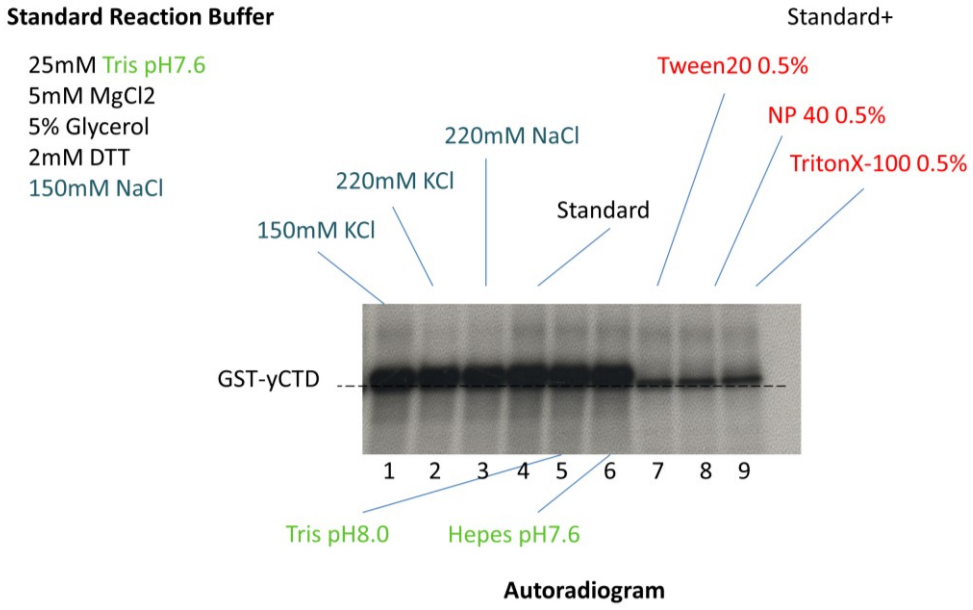

Figure 4.3: hCDK12/CyclinK complex CTD kinase activity is highly sensitive to detergents. Autoradiograms of SDS-PAGE gels of hCDK12/CyclinK kinase assays using GST-yCTD as substrate. Reaction conditions are 1 ug of GST-yCTD, 45min reaction time, 37 °C, and 30 µM ATP in 25 µL total reaction volume. Standard reaction buffer composition is given and the position of unphosphorylated GST-yCTD in the SDS-PAGE gel is indicated via a dashed line. Lane one contains the standard reaction conditions, while lanes 2-6 modify the salt, salt concentration, and buffering agent composition/pH. Lanes 7-9 represent the standard reaction buffer supplemented with the various detergents at the indicated concentrations.

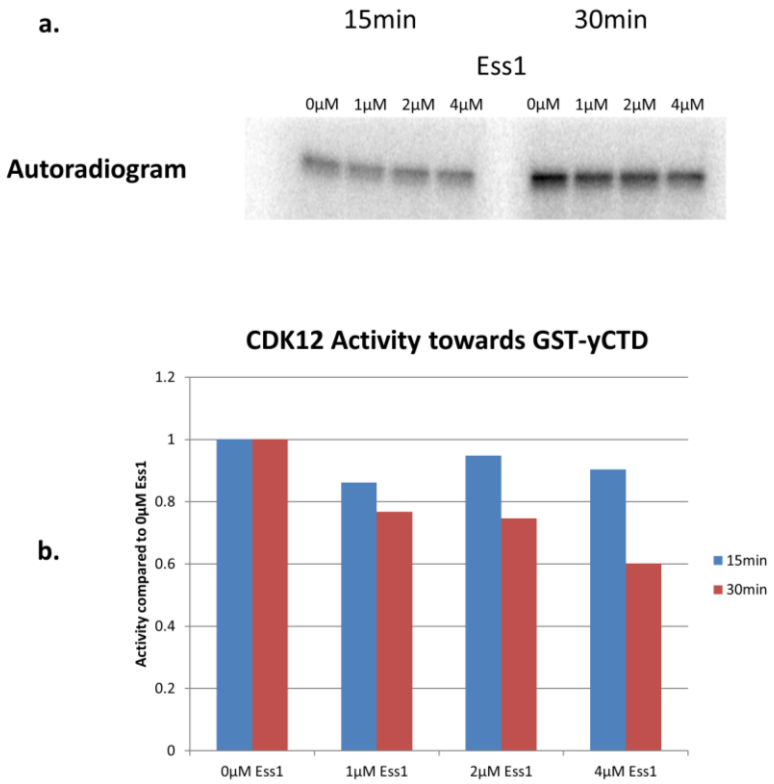

Figure 4.4: The effect of a prolyl isomerase on the activity of hCDK12. (a.) A representative autoradiogram of a SDS-PAGE gel from an hCDK12/CyclinK kinase assay with increasing amounts of the prolyl isomerase Ess1. Assays were performed using 1 ug of GST-yCTD as a substrate at 30 μM ATP, 37 °C, and 15/30 min reaction times. (b.) Quantification of the degree of phosphorylation of each mutant substrate using a PhoshpoImager and ImageQuant software (normalized to a value of 1.0 for the 0 uM Ess1). No stimulation of hCDK12 activity was observed.

detect any difference in GST-yCTD kinase activity between the full length and shortened CDK12/CyclinK complex (Fig. 4.2 d.).

In the interest of creating a negative control for our assays we also set out to generate a kinase dead mutant of CDK12. Based on the literature we focused our attention on mutating the conserved aspartate residue known to be essential for the phospho-transfer reaction in other CDKs (examples include: D145 of CDK2 and D167 of CDK9) (Garriga et al., 1996; van den Heuvel and Harlow, 1993); alignment of the

sequences of several CDKs identified the most probable candidate residue to be aspartate 877 of CDK12. Mutation of this residue to asparagine drastically reduced kinase activity to almost background levels, although some residual activity was observed on longer exposures and with high amounts of enzyme (Fig. 4.2 d., and data not shown). Following up on this result, we eliminated the possibility of a copurifying CTD kinase activity by purifying and assaying the D867N mutant using antibody saturated beads (size exclusion fractions were used as input and beads were subsequently washed with 1 M NaCl and 0.5% NP40); however the slight residual activity persisted. In view of this we made an additional mutation in the nucleotide binding site of CDK12: K756A modelled on the CDK2 K33A mutation (Leroy et al., 1996). The double mutant kinase had even less background activity, although some slight residual activity could still be detected at extremes of exposure and enzyme amounts (Fig. 4.2 d. and data not shown). Despite this slight residual activity, compared to wild-type kinase, for most intents and purposes both the single and double mutants can be considered to be inactive (also see (Joshi et al., 2014)). Accordingly, mutation of D877N and D877N/K756A inhibits the vast majority of kinase activity and confirms the purity of our CDK12/CyclinK purification.

In order to inform future experiments we also engineered an analog sensitive mutant of CDK12. Analog sensitive kinases are created by mutating a large phenylalanine residue, termed the "gatekeeper" residue, near the enzyme active site to a

much smaller glycine; the mutated enzyme is then able to accept bulky adenine analogues in its active site, allowing for specific inhibition of its activity (Blethrow et al., 2004; Shah et al., 1997). Alignment of CDK12 with other CDKs and comparison to other analog sensitive kinases (Larochelle et al., 2012) implicated Phenylalanine 813 of CDK12 as a candidate "gatekeeper" residue (Fig. 4.5 a.); this residue was mutated to glycine to create CDK12^{as}. Purified CDK12^{as} /CyclinK kinase activity was assayed alongside wild-type kinase in the presence of increasing amounts of the bulky adenine analog 1-NM-PP1. Addition of 1-NM-PP1 had no effect on the activity of the wild-type kinase in the GST-yCTD kinase assay; however, addition of 1-NM-PP1 to reactions containing CDK12^{as} resulted in a virtually complete loss of enzyme activity (Fig. 4.5 b.). Consequently, the two step purification process results in highly purified CDK12/CyclinK devoid of contaminating kinase activities. In addition, the high sensitivity of CDK12^{as} to 1-NM-PP1 suggests that it will be a useful tool for future studies of CDK12 kinase activity *in vivo*.

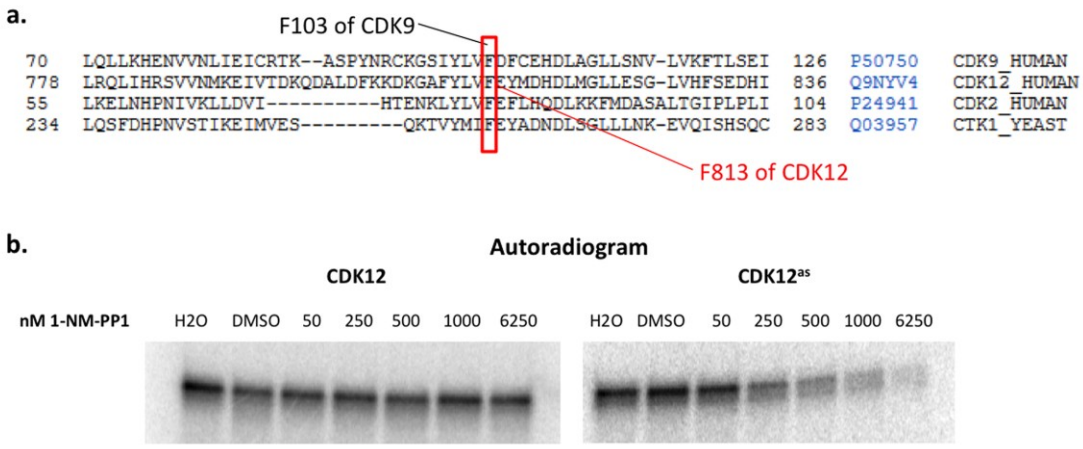

Figure 4.5: Developing an analog sensitive mutant of CDK12. (a.) Sequence alignment (UniProt [www.uniprot.org]) of CDKs (yeast Ctk1 and human CDK2, CDK9, and CDK12) identifying the potential "gatekeeper" residue of hCDK12. Relative amino acid positions are indicated at the beginning and end of each sequence and each protein's UniProt accession number and name are displayed on the right. Conserved phenylalanine residues that have been previously mutated to make analogue sensitive kinases are highlighted with a red box. (b.) Purified hCDK12/CyclinK (CDK12) and F813G hCDK12/CyclinK mutant (CDK12^{as}) CTD kinase activity was assayed using 1 ug of GST-yCTD fusion protein at 30 μM ATP, 37 °C, and a 30 min reaction time with increasing concentrations of the bulky adenine analogue 1-NM-PP1 in DMSO; the reactions were analyzed by SDS-PAGE and visualized using a PhosphorImager.

4.3.3 CDK12 substrate specificity and activity in the presence of DRB and FVP

Using a series of synthetic three heptad repeat phospho-peptides and truncated forms of the hCDK12/CyclinK complex, previous studies have reported that hCDK12 exhibits a marked preference towards a CTD phosphorylated at the Ser7 position within the heptad repeat (Bosken et al., 2014). In order to investigate the substrate specificity of the full length hCDK12/CyclinK complex towards a longer CTD substrate, we used bacterially expressed GST-CTD fusion proteins carrying ~15 wild-type or Ser-substituted repeats: WT (YSPTSPS), S2A (YAPTSPS), S2E, S5A, S5E, and S7E. These constructs either eliminate a potential phosphoacceptor (serine to alanine) or act as a

phosphomimetic (serine to glutamate, which resembles phospho-serine). Under our standard kinase assay conditions (300 μM ATP, 1 μg GST-CTD fusion protein substrate, 30 min reaction time), we find that hCDK12/CyclinK does exhibit a preference towards the S7E GST-CTD construct when compared to the other mutant CTD constructs, but not when compared to wild-type CTD construct which was phosphorylated to a similar

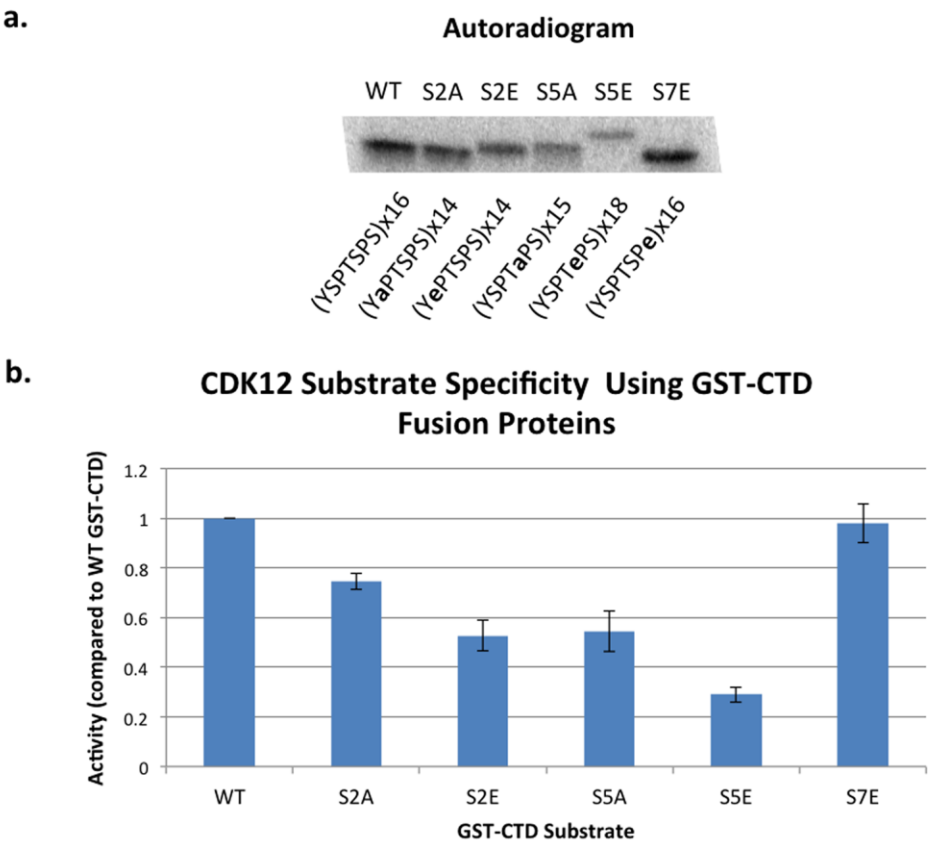

Figure 4.6: Phosphorylation of mutant GST-CTD substrates by hCDK12/CyclinK. (a.) A representative autoradiogram of a SDS-PAGE gel from an hCDK12/CyclinK kinase assay using mutant GST-CTD substrates. The composition and number of the CTD heptad repeats in the GST-CTD fusion protein is indicated below the appropriate lanes. (b.) Quantification of the degree of phosphorylation of each mutant substrate using a PhoshpoImager and ImageQuant software (normalized to a value of 1.0 for the wild-type GST-CTD fusion protein). Error bars are +/- 2 standard errors of the mean, n=3.

extent (Fig. 4.6 a. and b.). Kinase activity is significantly weaker towards the S2E and S5A substrates (~50% as compared to WT) with the S5E presenting itself as the worst substrate (~30% activity as compared to WT). The S2A construct was also a somewhat weaker substrate than WT or S7E (~75% activity as compared to WT) (Fig. 4.6 a. and b.).

Many studies of CDK9's role in RNAPII transcription and CTD phosphorylation employed the CDK9 inhibitor flavopiridol (FVP) (Chao and Price, 2001). FVP has been shown to exhibit significant selectivity for CDK9 as compared to the other transcriptional CDKs (IC50 of 3-20 nM as compared to ~300 nM for CDK7) and because of this presumed selectivity, the effects of FVP treatment have often been interpreted to be due to inhibition of CDK9 alone. However several other non-transcriptional CDKs (CDK1, 2, 4, and 6) are highly sensitive to FVP (IC50s at ~40 nM) (Schmerwitz et al., 2011) and experiments using a dominant negative form of CDK9 have suggested that the potent effects of FVP on transcription may be due to the inhibition of additional CTD kinases other than FVP (Garriga et al., 2010). Therefore, we set out to determine if full length CDK12 also exhibited sensitivity to FVP. The binding of FVP to CDK9 is characterized by a very strong, long lived, hydrophobic interaction, resulting in a one to one binding of inhibitor to kinase (even at nanomolar concentrations of inhibitor) (Baumli et al., 2008; Chao and Price, 2001). We decided to leverage this non-canonical inhibitor-kinase interaction, coupled with the more general behavior of the competitive

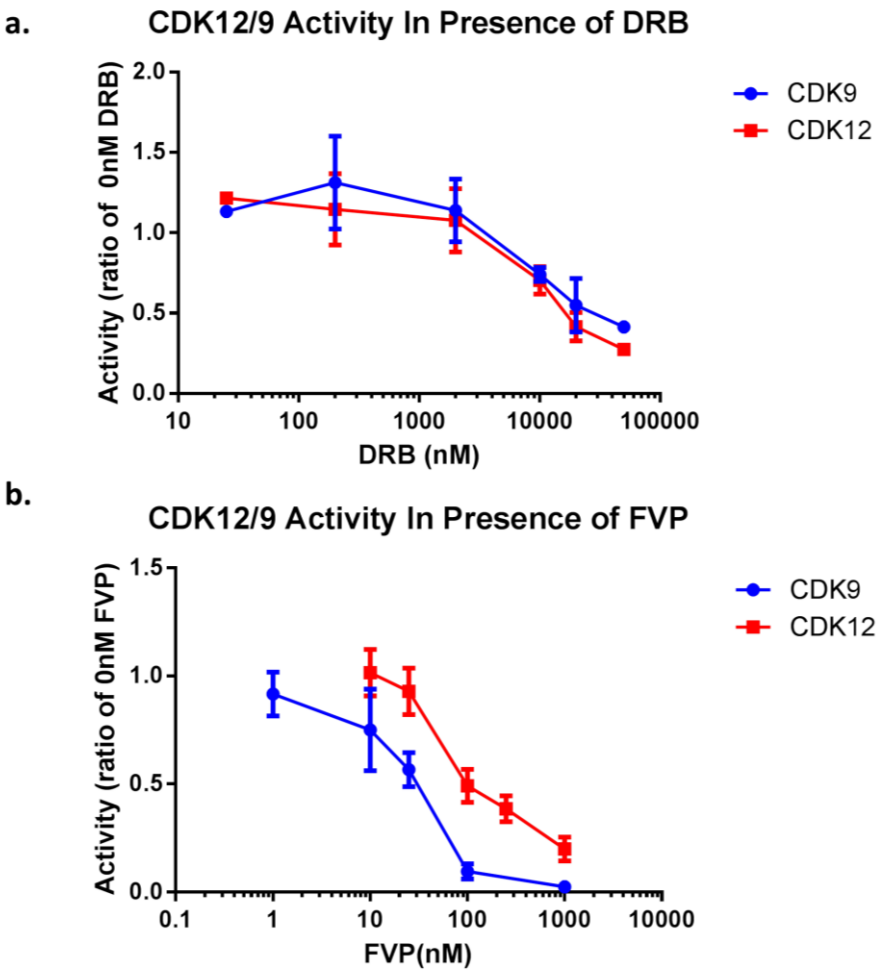

Figure 4.7: CDK9 and CDK12 activity in the presence of 5,6-Dichlorobenzimidazole Riboside (DRB) and Flavopiridol (FVP). (a.) Concentration series of (a.) DRB and (b.) FVP for CDK12/CycK and CDK9/CycT1 activity assayed using 1 ug of GST-yCTD (30 μM ATP, 37 °C, 30 min reaction time). Activity of each kinase is normalized to a value of one for 0 nM DRB or 0 nM FVP respectively. Error bars are +/- 1 standard error of the mean, n=3 (except for DRB 25 nM and 5000 nM, where n=1 and 2 respectively)

CDK inhibitor 5,6-Dichlorobenzimidazole riboside (DRB) in order to determine if CDK12 exhibits a similar susceptibility to inhibition by FVP. Assayed using GST-yCTD, under the same buffer conditions and activity, both CDK12 and CDK9 responded in the same way to increasing concentrations of the competitive inhibitor DRB (Fig. 4.7 a.).

Table 1: IC50 calculations for CDK9 and CDK12 activity in the presence of flavopiridol (FVP). IC50 values were determined using GraphPad Prizm 6.

	CDK9	CDK12
log(inhibitor) vs. normalized response		
Best-fit values		
IC50	26.03	147.6
95% Confidence Intervals		
IC50	18.01 to 37.62	104.6 to 208.2
Goodness of Fit		
Degrees of Freedom	14	14
R square	0.9180	0.9010
Absolute Sum of Squares	1628	1553
Sy.x	10.78	10.53
Number of points		
Analyzed	15	15

However, the behavior of CDK12 and CDK9 diverged when treated with FVP (Fig. 4.7 b.): Under the conditions used, CDK9 exhibited an IC50 of ~25 nM versus an IC50 of ~150 nM for CDK12 (Table 1). This suggests that the inhibition of CDK12 by FVP does not result in the tight interaction mode exhibited by FVP and CDK9. Despite this difference, the activity of CDK12 is clearly inhibited by the higher concentrations of FVP; therefore whether CDK12 is inhibited by FVP *in vivo* will depend on intracellular inhibitor and ATP concentrations and remains an open question.

4.4 Discussion

In order to further characterize CDK12's roles *in vitro*, and to inform future experiments *in vivo*, we have purified to near homogeneity full-length, active, human CDK12/CyclinK, the RNAPII CTD kinase orthologous to yeast CTDK-I. It is highly

likely that the key to heterologous expression of the full length CDK12 protein in the baculovirus/Sf9 expression system is the one to one expression of both the kinase and its cyclin partner. With this in mind, we used a "2A" peptide-linked multi-cistronic construct in order to co-express both hCDK12/CyclinK subunits and overcome previously reported solubility and post-translational modification problems. The over-expressed enzyme is heavily post-translationally modified, definitely by phosphorylation (probably on its N and C terminal extensions), but also through other modifications as phosphatase treatment does not completely resolve the diffuse nature of the hCDK12 band in SDS-PAGE gels. Although we do not know the exact degree of activating T-loop phosphorylation (a characteristic modification of CDKs that enables kinase activity) of the recombinant enzyme complex, the purified enzyme clearly exhibits CTD kinase activity towards an unmodified substrate (GST-yCTD); it would be interesting to see if co-expression of a CDK activating kinase (such as yeast CAK) further enhanced this activity. Fortuitously, due to its uncharacteristically large size, the purification of the recombinant His tagged hCDK12/CyclinK protein can be performed from nuclear extracts with a straightforward two step procedure yielding highly purified complex.

Enzymatic characterizations of the purified kinase *in vitro* reveal features important to understanding CTD phosphorylation *in vivo*. We find that the activity of hCDK12/CyclinK towards a GST-yCTD substrate is independent of the proline-rich

region C-terminal to the cyclin box of CyclinK and, unlike the activity of some CTD phosphatases (Werner-Allen et al., 2011; Xiang et al., 2010), is not stimulated by proline isomerization. Additionally, we observe that CDK12 kinase and CDK9 kinase (P-TEFb) show identical sensitivities to the inhibitor DRB and similar sensitivities to the inhibitor flavopiridol (FVP); consequently, when either DRB or FVP is used to inhibit CDK9 *in vivo*, it is likely that CDK12 is also inhibited. Therefore, attributing a particular event to either one of these kinases based on inhibitor studies is not straightforward and the studies that have done so merit a degree of re-evaluation.

We have also investigated the substrate specificities of full-length hCDK12 using GST-CTD fusion protein substrates with nominally 15 heptad repeats which are all WT (YSPTSPS) or are all altered at one position (S2A, S2E, S5A, S5E, S7E). The WT and S7E substituted CTDs are phosphorylated to fairly similar extents by hCDK12/CyclinK, while the S2A, S2E, S5A, and S5E present themselves as weaker substrates (75%, 50%, 50%, and 25% activity respectively). These data argue that *in vitro* hCDK12/CyclinK, much like CTDK-I (Jones et al., 2004) and P-TEFb (Czudnochowski et al., 2012), is a promiscuous kinase that can phosphorylate multiple serine positions within the CTD (as an example, with the S2A substrate the complex must phosphorylate either the S7 and/or S5 position within the heptad repeats). Although our data agree with the previously reported preference towards a substrate prephosphorylated at the Ser7 position of the CTD (Bosken et al., 2014) as compared to that of Ser2 or Ser5, we do not

observe this preference when comparing S7E to WT. The reason for this difference could be due to the nature of the substrate (our substrate is more reminiscent of the wild-type CTD in the number of heptad repeats [16 as compared to 3 in the other study], however, the use of glutamate as a phosphomimetic is inferior to a true phospho-serine) or due to the use of a full length hCDK12/CyclinK complex. Regardless, the preference of CDK12 for S7P and unphosphorylated repeats is interesting; it is likely to be indicative of the phospho-CTD patterns recognized by CDK12 *in vivo* and, with further study, may play a role in furthering our understanding of the phospho-patterning of the CTD.

Finally, as a tool for future studies, we have engineered, expressed, and purified an analog sensitive form of hCDK12/CyclinK. Not all kinases are amenable to the creation of analog sensitive mutants; therefore we were very excited to find that the mutated form of CDK12 was stable and highly sensitive to the cell permeable, bulky adenine analog 1-NM-PP1. We anticipate that the availability of full length enzyme and its analog sensitive mutant will be useful in teasing apart the structural and catalytic roles of hCDK12/CyclinK (allowing for comparisons between total CDK12 depletion/knockout versus kinase activity inhibition) and will lead to an improved understanding of hCDK12/CyclinK activities within the cell and CTD modification and RNAPII transcription in general.

5. Insertion of a CDK12 Analog Sensitive Mutation into a Human Cell Line

5.1 Introduction

Since our initial characterization of CDK12/CyclinK as a Ser2 CTD kinase the structure of its kinase homology domain has been elucidated, it has been shown to play a role in the 3′ end processing of the MYC gene and it has been identified as a tumor suppressor for ovarian cancer (Bosken et al., 2014; Davidson et al., 2014; Joshi et al., 2014). Most intriguingly, and particularly relevant to its role as a tumor suppressor, CDK12/CyclinK has been implicated in the maintenance of genomic stability through the regulation of DNA damage response genes and as a determinant of PARP1/2 inhibitor sensitivity in the treatment of cancer (Bajrami et al., 2014; Blazek, 2012; Blazek et al., 2011). Despite this progress there is still much that is not known about the *in vivo* roles of CDK12/CyclinK at the molecular level. As an example, its role in the regulation of DNA damage response genes has not been fully elucidated: Is the downregulation of these mRNAs following CDK12 depletion a result of the polymerase occupancy changes at the 5′ end of the genes as posited by Blazek et al., or is it actually due to errors in mRNA processing or a consequence of long-term secondary effects? Given our list of dCDK12 and hCDK12 associated proteins (see chapter 3), it is likely that CDK12/CyclinK plays both catalytic and noncatalytic roles in mRNA processing and other transcription related functions, further complicating analysis. Because there is no specific hCDK12 inhibitor, and because RNAi mediated depletion affects both the

structural and catalytic elements of CDK12 function, gaining a better understanding of CDK12's role in transcription will be challenging without the development of additional molecular tools. Having shown that an analog sensitive form of hCDK12/CyclinK (CDK12^{as}) is stable, highly sensitive to the adenine analog 1-NM-PP1, and exhibits wild-type levels of activity prior to inhibition *in vitro*, we wanted to come full circle and reintroduce CDK12^{as} into a human cell line. Therefore, the recent discovery and utilization of the CRISPR/Cas9 system as a genome engineering tool presented itself as an exciting opportunity for the creation of a powerful tool for future experiments (Mali et al., 2013; Ran et al., 2013).

In its natural state CRISPR/Cas is a microbial immune system which employs a RNA-guided nuclease in order to cleave foreign genetic elements. The system's name comes from an array of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), which are separated by short variable sequences from exogenous DNA targets called protospacers. Each protospacer is associated with a Protospacer Adjacent Motif (PAM) which varies according to the specific CRISPR system in question: In the type II CRISPR system, which consists of the Cas9 nuclease, a CRISPR RNA array that encodes a guide RNA and a trans-activating crRNA that facilitates the processing of the array into discrete units, the target DNA must precede a 5' NGG PAM. Once it is transcribed and processed, each discrete unit of the CRISPR array contains a 20 nucleotide guide sequence which then directs Cas9 to a 20bp DNA target by Watson-

Crick base pairing. In mammalian cells this RNA-guided nuclease system can be reconstituted through the heterologous expression of human codon optimized Cas9 and the appropriate RNA components (the guideRNA and trans-activiating crRNA), which are fused together to create a chimeric single guide RNA (gRNA). Thus Cas9 can be targeted toward any genomic region of interest resulting in a site specific double strand break. Cleavage and repair of the DNA at the targeted site by nonhomologous end joining (NHEJ) can stochastically generate frameshifts in a targeted open reading frame (ORF) and result in knockout; or if an appropriate repair template is provided, homology-directed repair can generate specific gene modifications at the targeted loci (see protocol by Ran et al. for more details) (Ran et al., 2013). Therefore, in order to dissect the indirect and direct roles of CDK12 in RNA processing and other events, we set out to engineer CDK12 knockout and analog sensitive human cell lines using the CRISPR/Cas9 system.

5.2 Materials and methods

5.2.1 Knockout cell line creation

The Cas9 expression plasmid (pcDNA3_3 TOPO_CAS9) and gRNA expression vector (pCR_BluntII_TOPO_U6_gRNA) were obtained from Nicholas Barrows and Dr. Mariano Garcia-Blanco (originally ordered from Addgene; designed by the Church lab).

In order to create a knockout cell line a protocol similar to the one found in Ran et al. was followed (Ran et al., 2013). In detail: 3 guide RNAs were selected (from

computationally designed gRNAs in the Mali et al. supplemental materials (Mali et al., 2013) and at http://arep.med.harvard.edu/human_crispr/) to target exon 1 of hCDK12.

Guide RNA #1: GAGCTGCCGCCTCCCGATGACGG

Guide RNA #2: GCTTGTGCTTCGATACCAAGCGG

Guide RNA #4: GGCCTTCAAACTAGACCGAAGGG

Each guide RNA was ordered as a "top" and "bottom" segment designed to be annealed and extended (see below) in order to form a double stranded construct that could be integrated into the gRNA vector plasmid. The "top" segment consists of TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC followed by the guide RNA sequence (lacking the final 3 nucleotides representing the PAM site (NGG) as these are included in the vector). The "bottom" segment consists of GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC followed by the reverse complement of the 20 nt guide RNA sequence (again lacking the final 3 nucleotides). The final top and bottom segments are 60 bp in length and were ordered from IDT and are shown below (Table 2) with the extensions italicized and highlighted in blue. The pairs of guide RNA segments were resuspended at 100 μM, mixed, and annealed/extended using PWO Master (A 2x high fidelity PCR master mix; Roche). This was done by performing 5 cycles of PCR with a 55 °C annealing and a 30 second extension step. The pCR Blunt II TOPO U6 gRNA vector was linearized with AflII

Table 2: hCDK12 guide RNA segment sequences. gRNA expression vector sequence used for sequence independent ligation (Gibson reaction) is shown in blue.

Guide RNA #- Strand	Sequence
1-Top 1-Bottom	*5′-* *TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC*GAGCTGCCGCCTCCCGATG-3′ *5′-* *GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*TCATCGGGAGGCGGCAGCTC-3′
2-Top 2-Bottom	*5′-* *TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC*GCTTGTGCTTCGATACCAAG-3′ *5′-* *GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*CTTGGTATCGAAGCACAAGC-3′
4-Top 4-Bottom	*5′-* *TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC*GGCCTTCAAACTAGACCGAA-3′ *5′-* *GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*TTCGGTCTAGTTTGAAGGCC-3′

restriction endonuclease (NEB). The annealed gRNA insert and cut vector were purified using the QiaQuick PCR Purification Kit (Quiagen) and the products eluted in 30 µL dH_2O. The linearized gRNA vector was then diluted to 8.65 ng/µL and the 100 bp gRNA insert to 1.11 ng/µL. Gibson assembly was used to incorporate the gRNA insert into the gRNA vector; briefly 5 µL of diluted gRNA vector was added to 5 µL of diluted insert, then 10 µL of 2x Gibson assembly master mix was added (NEB) and the reaction was incubated for 1 hour at 50 °C. The reaction mixture was then diluted 1:4 with dH_2O and 2 µL was used to transform XL-1 blue chemically competent cells. The transformation reactions were selected on Kanamycin and the resulting colonies were

screened for successful incorporation of the gRNA inserts by sequencing of plasmid DNA via the M13R primer.

HeLa cells were then transfected using 1.5 μg of pcDNA3_3 TOPO_CAS9 and 0.75 μg of two gRNA vectors (guideRNA 1 and 4; 1 and 2; and 2 and 4) in a 6 well plate format using Lipofectamine 2000 (Life Technologies), following the manufacturers protocol. 24 hours after transfection, the cells were passaged to T-75 flasks and placed under G418 (Gibco) selection (1 mg/mL; effective concentration of G418 was determined beforehand) for 4 days. Following selection, each transfected cell line was analyzed for Cas9 activity via genomic DNA purification (DNeasy Blood and Tissue Kit, Quiagen) followed by PCR using primers outside of the targeted region (Forward Primer: TTCTTCAGGTCAGGGGAAAG; Reverse Primer: CTGGTGGTGACGATGTTTGT). Single clones were then isolated by low density plating, propagated, and analyzed for the presence of hCDK12 via western blotting.

5.2.2 Analog sensitive cell line creation

In order to create an analog sensitive cell line another guide RNA targeting exon6 of hCDK12 was designed by locating a unique 20 bp sequence with the appropriate PAM motif (NGG) as near to the F813 codon as possible.

AS gRNA: GGTCCATATACTCAAATACAAGG

As in 5.2.1, two oligonucleotides were ordered from IDT (Table 3).

Table 3: hCDK12 analog sensitive mutant guide RNA segment sequences. gRNA expression vector sequence used for sequence independent ligation (Gibson reaction) is shown in blue.

Guide RNA #- Strand	Sequence
AS-Top AS-Bottom	*5′-* *TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC*GGTCCATATACTCAAA TACA-3′ *5′-* *GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*TGTATTTGAGTATATGG ACC-3′

The final AS gRNA vector was produced as detailed for the other constructs and HeLa cells were transfected with the Cas9 and AS gRNA plasmids as detailed above (see 5.2.1). Cas9 activity and targeting were confirmed through isolation of genomic DNA and the Surveyor Assay (Transgenomics) (For details refer to (Ran et al., 2013) and manufacturer's protocol; surveyor assay primers were as follow: Forward-AGAACCTGTGGTCAACTCCAA and Reverse – TCCCACACTGAACGTAGGAA, data not show).

200 ng of a double stranded 850 bp homologous repair template was ordered from IDT (gBlocks Gene Fragment service) containing ~425 bp of the genomic sequence flanking either side of the cut site with the analog sensitive mutation and a few modifications (Fig. 5.1).

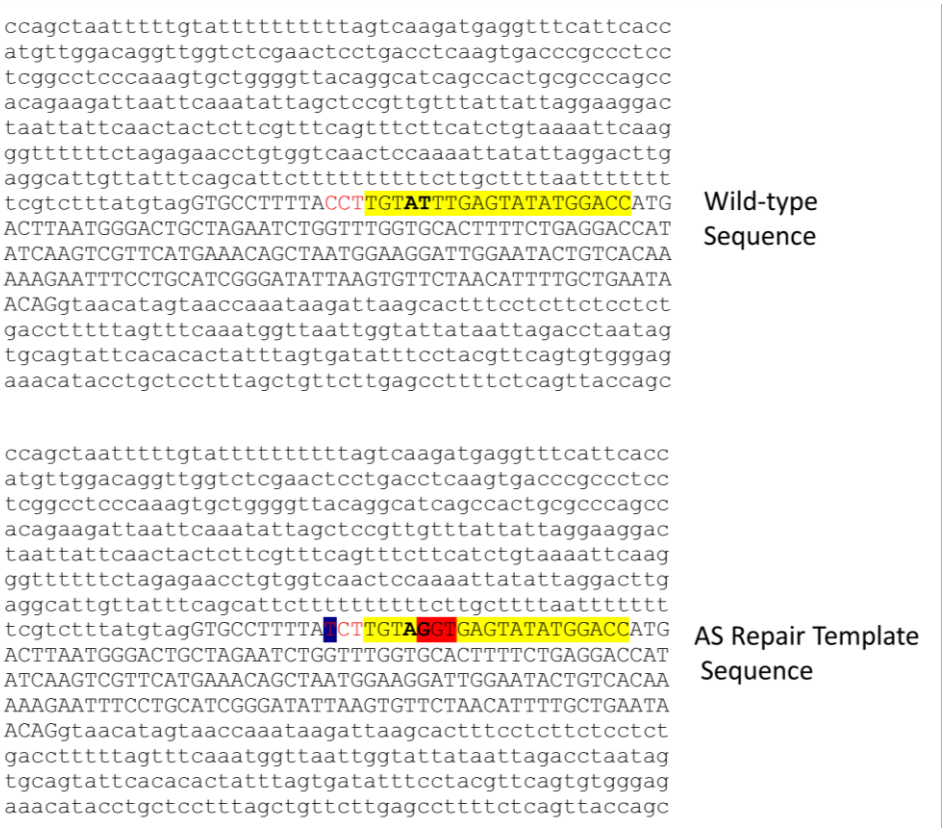

Figure 5.1: Sequence surrounding Exon6 of hCDK12 and the analog sensitive repair template. Genomic sequence was obtained using the UCSC genome browser (https://genome.ucsc.edu/). Exons are displayed in caps, introns in lowercase. Red letters are the PAM sequence (required for targeting of the Cas9 nuclease), the bolded AT is the Cas9 cut site. The WT sequence and analog sensitivity inducing repair template (shortened from its full 850bp for display purposes) is shown up top and below respectively. The gRNA complementary sequence used to target Cas9 to Exon6 is shown in yellow. The repair template includes a TTT to GGT (in red) which results in F813G. There is also a silent mutation in the PAM to ensure that Cas9 does not cut the repair template or the repaired sequence (C->T in blue).

HeLa cells were transfected with the Cas9 expression plasmid, AS gRNA expression plasmid, (as detailed above) and 50 ng of the repair template. Following selection, single clones were isolated by low density plating and propagated. Genomic DNA from each clone was assayed using PCR with forward primers specific for WT (GGTGCCTTTTACCTTGTATTTGA) and analog sensitive (GTGCCTTTTATCTTGTAGGTGAG) sequence and a reverse primer outside the region

encompassed by the repair template (CCAATGCAGTATAGGCAAATAAA). A single clone was found which did not yield a WT PCR product, but did yield an analog sensitive PCR product. The region of interest was then PCR'd from the genomic DNA of this cell line (Forward Primer: AGAACCTGTGGTCAACTCCAA and Reverse Primer: CCAATGCAGTATAGGCAAATAAA) and sequenced (sequencing primer: ATAGAGCCGAGCAAGTCCAA). The resulting PCR product and TA clones of the product in pCR2.1-TOPO (Topo TA cloning kit, Invitrogen) were then sequenced (Duke DNA analysis facility)

5.2.3 Antibodies and qPCR

rAP anti-hCDK12: Affinity-purified IgGs directed against a peptide comprising amino acids 201-220 of hCDK12 (NCBI RefSeq: NP_057591.2).

RNA was isolated using the RNeasy mini kit (Quiagen) (The optional on-column DNAse step was performed) and cDNA synthesis was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) using 1.5 μg of total RNA as input. Quantitative PCR (qPCR) was performed using Power SYBR Green PCR master mix (Applied Biosystems) and the StepOnePlus real-time PCR system (Applied Biosystems); each sample was normalized based on the amount of β-Actin and changes in gene expression were calculated using the ΔΔCt method. The forward and reverse primers used to detect each target are listed below.

CDK12: AACACTGATGGGCCTGAAAC and GTTCTTCACCAGGGTCTGGA

CyclinK: AAACCTGGACCACACAAAGC and GCCCACATCAAAGATGAACC

β-actin: GCTCGTCGTCGACAACGGCTC and CCTCGTCGCCCACATAGGAATC

5.3 Results

5.3.1 Constructing a CDK12 knockout cell line

As a true CDK12 knockout cell line was likely to exhibit informative phenotypes and be very useful for rescue experiments, we began our foray into gene engineering by attempting to delete CDK12 through the isolation of a clone harboring a Cas9 induced frameshift mutation. HeLa cells were transfected with Cas9 expression vector and two guide RNA (gRNA) expression vectors (gRNA #1 & #4, #1 & #2, and #2 & #4; see 5.2.2 for details). The use of two separate gRNA vectors targeting two closely spaced regions can result in a deletion of the sequence separating the two sites (through NHEJ), allowing

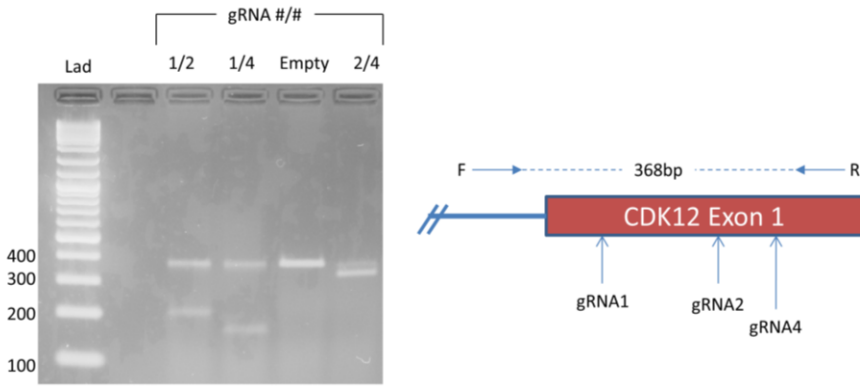

Figure 5.2: Analysis of gRNA targeting and Cas9 activity via PCR. A schematic of hCDK12 exon 1 with the gRNA target sites and PCR primer positions indicated, is displayed on the right; the PCR product from wild type HeLa cells is expected to be 368 bp. PCR results from pooled genomic DNA of Cas9 and gRNA transfected cells are visualized on the left using a 1.5% agarose gel stained with EtBr. HeLa cells were transfected with the indicated pairs of gRNA vectors (Empty is an empty vector control). PCR is performed on pooled genomic DNA from each transfection; note the appearance of NHEJ mediated deletions in the 1/2, 1/4, and 2/4 lanes.

for an easy way to assess if Cas9 is correctly targeted and active, via PCR. Following selection and propagation, total genomic DNA was isolated from the pools of transfected cells and analyzed; all 3 of the gRNA constructs were found to be active (Fig. 5.2). Single clones were then isolated by low density plating, propagated, and analyzed for the presence of hCDK12 via western blotting. Unfortunately, attempts at creating CDK12 knockout HeLa cells by targeting the 5′ end of the CDK12 ORF were not successful; we were able to isolate several clones that exhibited truncations due to NHEJ events (even in cells transfected with a single gRNA), but no frameshift mediated null clones were isolated (~40 clones were assayed in total) (Fig 5.3). This may be due to the essential nature of CDK12; although our failure to isolate a null clone does not prove that CDK12 is necessary for viability in HeLa cells, it does agree with previously

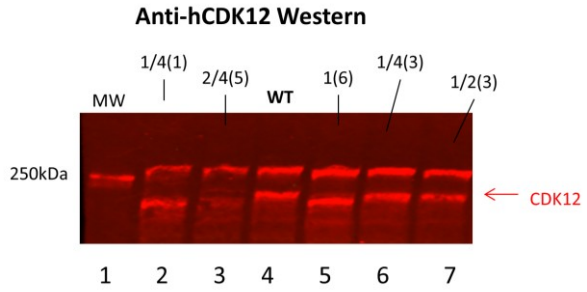

Figure 5.3: hCDK12 westerns of whole cell extracts from putative CDK12 knockout HeLa cell clones. Western blotting with an anti-hCDK12 antibody reveals the presence of truncated forms of hCDK12 protein in several clones targeted for CDK12 knockout via CRISPR-Cas9. The top band (above the 250 kDa molecular weight marker (lane1)) is an off target, cytoplasmic protein recognized by the anti-CDK12 antibody (also see Fig. 5.6). Compare the molecular weight of the CDK12 signal from the clones in lanes 2, 5, and 3 to WT HeLa cells (lane 4). Clone 2/4(5) (lane3) was subsequently re-assayed and found to contain a shortened hCDK12 protein present at about 50% of the wild type amounts (a likely heterozygote with a null at one allele).

reported data using CDK12 knockout mice and Drosophila (Blazek et al., 2011; Rodrigues et al., 2012) which found the gene to be essential for development.

5.3.2 Constructing an analog sensitive CDK12 cell line

In order to create an analog sensitive CDK12 cell line HeLa cells were transfected with the Cas9 expression vector, a guide RNA expression vector targeting the genomic position of P813 of CDK12, and a repair template harboring a mutation in the PAM (to avoid Cas9 targeting to the repair template and repaired regions) and the analog

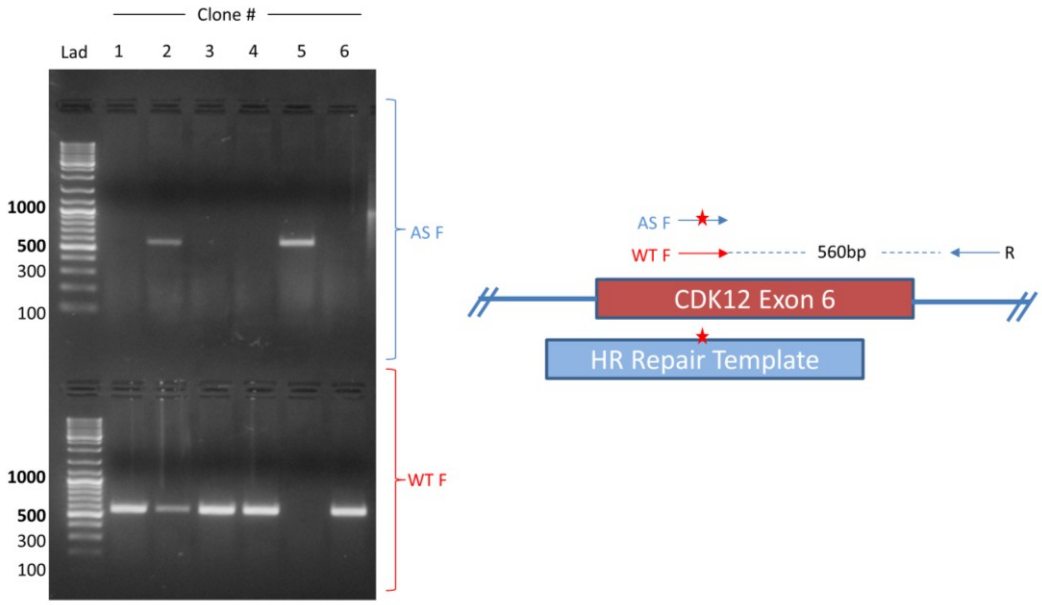

Figure 5.4: Analysis of putative analog sensitive clones via PCR. A schematic of hCDK12 exon 6 with the positions of the homologous repair template (blue box), the analog sensitive mutation (red star), and the PCR primers specific to the analog sensitive (AS) and wild type (WT) sequences indicated is displayed on the right; the PCR product is expected to be 560 bp. PCR of genomic DNA from 6 putative analog sensitive clones is visualized on the left using a 1.5% agarose gel stained with EtBr; the top section of the gel displays the results using AS specific primers while the bottom section of the gel displays the results using the WT primers. Note clone #5, which appears to be homozygous for the analog sensitive mutation.

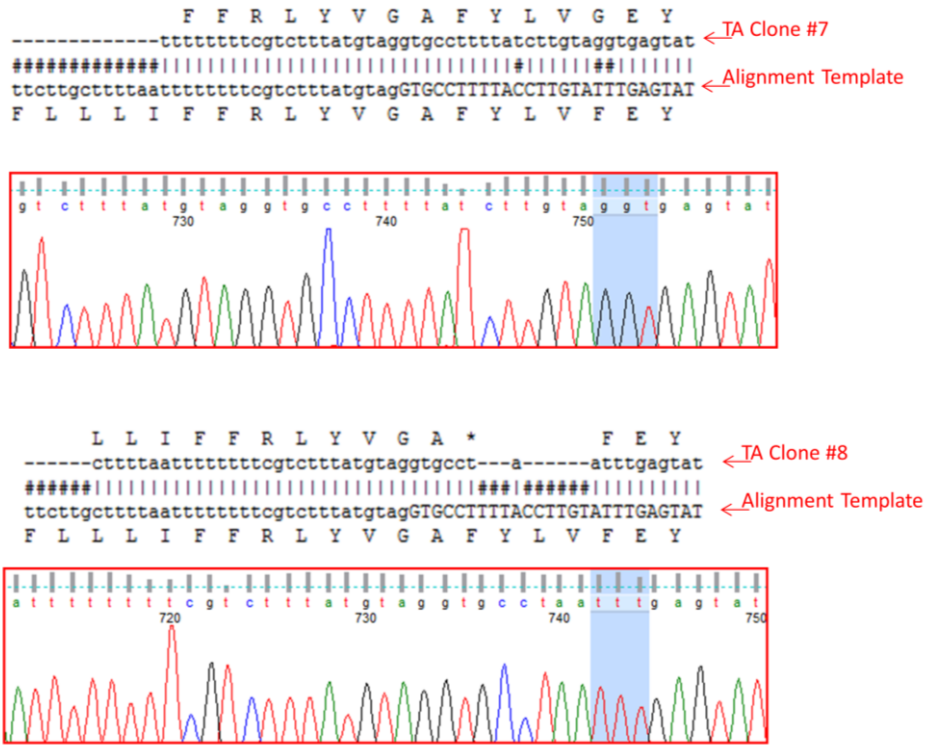

Figure 5.5: Sequences of hCDK12 exon 6 genomic DNA from the putative CDK12 analog sensitive cell line. Plasmids containing TA cloned PCR fragments from hCDK12 genomic DNA surrounding exon 6 were submitted for Sanger sequencing. Alignment to the canonical sequence, the translated protein sequence of the ORF, and chromatographs are shown; in the alignment template exons are displayed in upper case and introns in lower case; in the chromatographs the codon for F/G 813 is highlighted in blue. 8 TA clones were analyzed and only the two displayed sequences were observed (6 analogue sensitive and 2 wild-type with the premature stop codon). TA clone #7 (top) exemplifies the analogue sensitive allele, TA clone #8 exemplifies the truncated wild-type allele (The * in the translation indicates a stop codon).

sensitive mutation P813G (see Figure 5.1 and section 5.2.2). Selection, followed by the isolation of single colonies by low density plating and propagation, led to the identification of a single clone which appeared to be homozygous for CDK12^{as} via PCR using analog sensitive and wild type specific primers (Figure 5.4).

Sequencing of genomic DNA confirmed this putative CDK12^{as} cell line to be a heterozygote consisting of two alleles, one harboring a NHEJ induced stop codon in the

targeted position and one containing a clean analog sensitive mutation (Fig 5.5). This genotype was confirmed via western blotting of whole cell extracts for hCDK12 and through mRNA quantification by qPCR which showed ~50% decreased levels of CDK12 and CDK12 mRNA as compared to the parental cell line (the premature stop codon induces nonsense mediated decay of the associated mRNA) (Fig. 5.6).

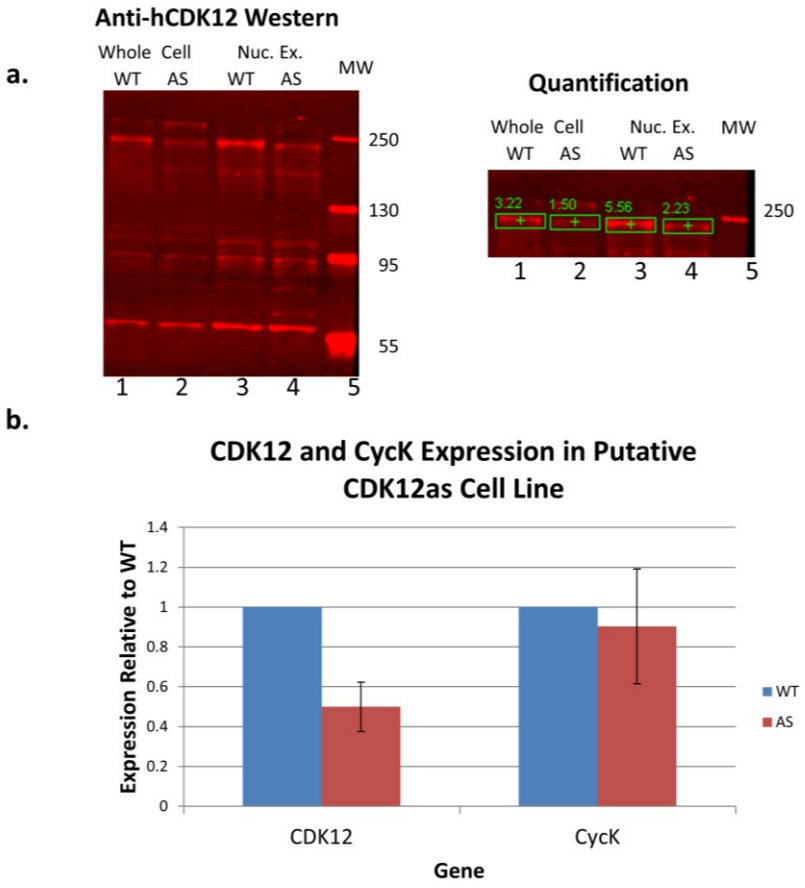

Figure 5.6: hCDK12 expression in the putative CDK12 analog sensitive cell line. (a.) hCDK12 western blot on whole cell (lanes 1-2) and nuclear (lanes 3-4) extracts from the parental (WT) and putative CDK12 analog sensitive (AS) cell line. The amount of hCDK12 in the analog sensitive cells was quantified to be ~50% to that of the parental cell line (on the right; Licor Odyssey software). The anti-hCDK12 antibody targets the N-terminal arm of hCDK12 (amino acids 201-220), well before the stop codon at amino acid 813 (b.) mRNA quantification of hCDK12 and CyclinK mRNA levels using qPCR. Expression values are standardized to the parental cell line (WT) whose value was set at 1, error bars are +/-2 standard errors of the mean, n=3.

Unfortunately, kinase assays of immunoprecipitated CDK12 isolated from the engineered cell line (data not shown) and sequencing of hCDK12 cDNA revealed an unforeseen problem. Mutation of the P813 codon to glycine results in the formation of an alternative splice acceptor site, which was now employed by the mutant cell line resulting in a protein containing an 18 amino acid deletion (Fig. 5.7); this protein exhibits

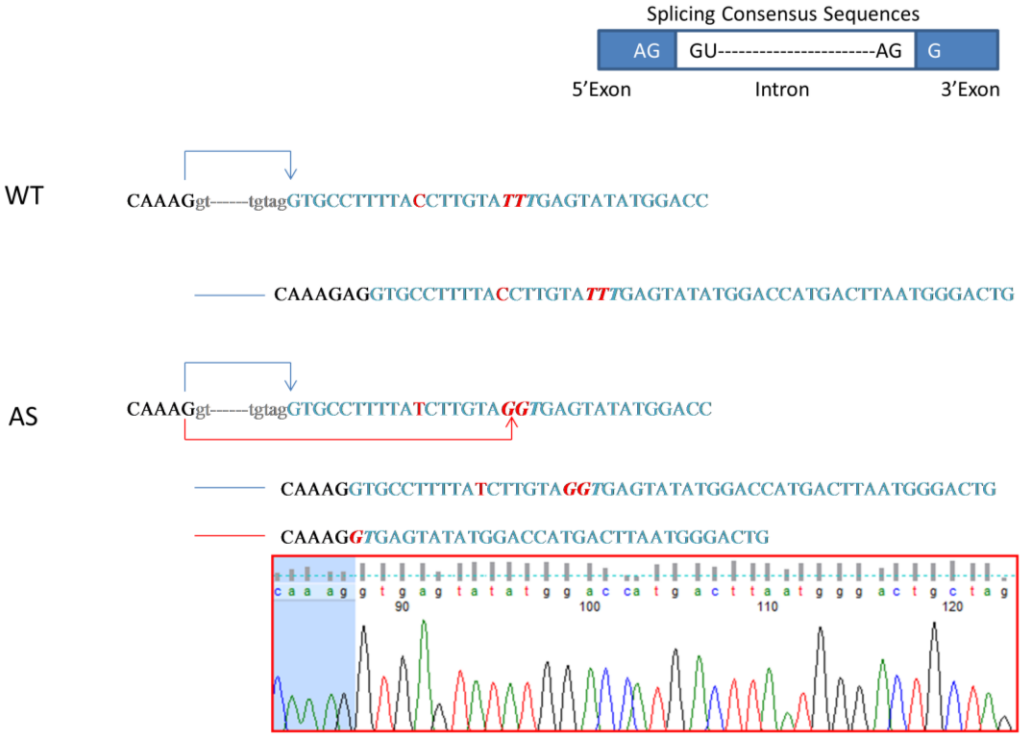

Figure 5.7: Schematic of splicing issues in the putative hCDK12 analog sensitive cell line. Mutation of the TTT codon encoding P813 of hCDK12 to GGT (P813G) results in the creation of a novel splice acceptor site downstream of the canonical one between exon5 and exon6 of hCDK12. Consensus splicing sequences are shown at the upper right of the figure. A schematic of exon 5 (black text), the intron between exons 5 and 6 (gray lowercase text), and exon 6 (blue text) is depicted in the parental (WT) and putative analog sensitive cell line (AS). The PAM and analog sensitive mutations are highlighted in red. Note how mutation of P813G with TTT->GGT results in the formation of a "tgtag" sequence motif exactly the same as the motif present at the end of the intron. Chromatographs of TA cloned cDNA sequences from this site are shown below the mis-spliced sequence; the end of exon 5 is highlighted in blue. This unanticipated splicing results in an 18AA truncation of the hCDK12 protein; the resultant protein retains residual activity, but is not analog sensitive (data not shown). As all codons encoding glycine begin with GG, a silent mutation in the preceding codon could potentially rectify this problem.

residual kinase activity, but is not analog sensitive. We are currently working to determine whether further modification of the alternative splice acceptor site or the nuclease targeting site mutation will resolve this issue. Therefore, construction of a CDK12 analog sensitive cell line using the CRISPR-Cas9 system is likely to be a lucrative endeavor; however, simply inserting the analog sensitive mutation into the genome can perturb the splicing of CDK12 exon 5 and 6.

5.4 Discussion

In order to create a set of tools to further study hCDK12/CyclinK's role in transcription and to test the proposals posited in chapter 3, we set out to use the CRISPR-Cas9 system to engineer knockout and analog sensitive CDK12 HeLa cells. Although we were unable to isolate a knockout clone of hCDK12, likely due to the essential nature of the kinase, we have demonstrated the feasibility of creating analog sensitive hCDK12 cell lines using this approach. Unfortunately, the verification of our haploid CDK12^{as} cell line failed at the final stages due to the creation of an unforeseen splice acceptor site at the position of the F813G mutation. We hope that our problems with the splicing of the analog sensitive hCDK12 cell line are a useful cautionary tale to others attempting similar experiments and aim to complete the construction of a true analog sensitive cell line in the near future. We anticipate that this tool will be useful in teasing apart the structural and catalytic roles of hCDK12/CyclinK and will lead to an

improved understanding of hCDK12/CyclinK activities within the cell and CTD modification and RNAPII transcription in general.

6. Conclusions and Future Directions

6.1 Conclusions

The primary aim of this dissertation was to explore the comparative genomics-based hypothesis that the previously-unstudied *Drosophila* CDK12 (dCDK12) and little-studied human CDK12 and CDK13 (hCDK12 and hCDK13) proteins are RNAPII CTD elongation-phase kinases, the metazoan orthologs of yeast Ctk1. By combining the evolutionary and functional evidence from previous studies with a variety of experimental techniques, we have determined that both dCDK12 and hCDK12 are indeed responsible for proper CTD phosphorylation *in vivo*, and concluded that CDK12 is indeed a bona fide CTD kinase, the metazoan ortholog of yeast Ctk1; intriguingly the function of hCDK13 remains ambiguous. These findings have clarified the evolutionary and functional relationships between the two yeast CDKs, Ctk1 and Bur1, and their metazoan homologues; but most importantly they have drawn attention to major metazoan CTD kinase activities that have gone unrecognized and unstudied until now. The recognition of these activities allows for a better informed investigation of the links between cotranscriptional processes and RNAPII transcription in higher eukaryotes.

Over the course of the last few years, several of our initial expectations, which assumed that any process or event dependent on normal phosphorylation of the CTD on elongating RNAPII will be affected in cells deficient for CDK12, have been verified. Specifically, hCDK12 has been shown to be important for the 3′ end processing of the

MYC gene (Davidson et al., 2014) and dCDK12 has been shown to help mediate splicing decisions in *Drosophila* blood brain barrier differentiation (Rodrigues et al., 2012). In addition to splicing and 3' end processing, we also anticipate that other events such as transcription-coupled chromatin modification, snRNA termination, and mRNA export out of the nucleus will be affected by loss of CDK12 activity. However, it is worth noting that the extent of these effects is likely to be influenced by the specific transcripts and/or cell types in question: It has been shown that in *C. Elegans,* CDK12 mediated Ser2 CTD phosphorylation is highly enriched in the germline as compared to somatic cells (Bowman et al., 2013) and CDK12 expression is variable, with the highest mRNA levels present in testis and ovaries (Edwards et al., 1998; Ko et al., 2001). Despite such challenges we expect that the full extent of CDK12 mediated co-transcriptional processing will become apparent over the next few years.

In conjunction with its roles in mRNA processing, since our initial characterization, CDK12 has been implicated in two interesting phenomena. The first is the maintenance of self-renewal in embryonic stem cells, where depletion of CDK12 resulted in differentiation (Dai et al., 2012), and the second is in the maintenance of genomic stability (Blazek, 2012). In accordance with the latter of these activities, CDK12 has been identified as a tumor suppressor for ovarian cancer (Joshi et al., 2014) and as a determinant of PARP1/2 inhibitor sensitivity (Bajrami et al., 2014). Intriguingly, alterations in transcription-associated homologous recombination events have been

observed in Ctk1Δ yeast strains (Grenetier et al., 2006; Westmoreland et al., 2009), and certain proteins important to genome stability are linked to transcription via direct interaction with the hyperphosphorylated CTD (Islam et al., 2010; Kanagaraj et al., 2010; Winsor et al., 2013); mice missing one of these phospho-CTD associating genome stability proteins (RECQ5) exhibit an increased incidence of cancer (Hu et al., 2007). However, it is not known whether the effects of CDK12 depletion on genomic stability are mediated through the loss of direct phospho-CTD interactions by proteins involved in the DNA damage response or through the deregulation of transcription and mRNA processing (or both). Thus, we anticipate that further characterization of CDK12 and its functions will be of broad scientific interest and significant medical relevance.

Therefore, in order to further expand on our initial characterization and to facilitate future studies, we have also identified putative CDK12 associated proteins, expressed and purified the full-length active hCDK12/CyclinK complex, and attempted to design an inhibitable version of CDK12 for *in vitro* and *in vivo* studies. Our *in vitro* characterizations of the hCDK12/CyclinK complex offer additional insights into CDK12 substrate preference and highlight the issue of inhibitor specificity in previous studies. Most intriguingly, our identification of CDK12 associated proteins suggest that CDK12 affects RNA processing events in two distinct ways: Indirectly through generating factor-binding phospho-epitopes on the CTD of elongating RNAPII and directly through binding to specific factors. In terms of differentiating between these possibilities, we

hope that our design and characterization of an analog sensitive hCDK12 mutant will prove a powerful tool for future studies and that our splicing problem in the analog sensitive hCDK12 cell line is a useful cautionary tale to others attempting similar experiments.

Overall, this dissertation has helped to fill important gaps in our understanding of eukaryotic transcription, its associated processes, and the links between them. We have highlighted the metazoan Ser2 CTD kinases as an area ripe for further investigation and developed useful tools for teasing apart the structural and catalytic roles of CDK12. We hope that these discoveries will continue to lead to an improved understanding of hCDK12/CyclinK activities within the cell and CTD modification and RNAPII transcription in general.

6.2 Future directions

As discussed in previous sections, the characterization of the *in vivo* roles of CDK12 and CDK13 is still a work in progress, especially at the molecular level. In the long term, many important questions will need to be answered including, but not limited to the following: Where exactly, and when, do these kinases add phosphates along the CTD? Why are there two Ctk1 homologues in humans and what is the actual function of CDK13? To what extent do these kinases act at specific subsets of genes or in particular cell types? What is the actual mechanism behind CDK12 mediated genomic stability? How much of CDK12's function is due to its catalytic activity versus its

structural elements? Over the next several years, further study of CDK12 and CDK13 will likely answer these and other questions and should lead to many fascinating discoveries.

In terms of more immediate short-term goals, we believe that the creation of an analog sensitive human cell line would be an extremely powerful tool for answering many of the above questions. Having shown the efficacy of the analog sensitive mutation, we aim to overcome the splicing issues discussed in section 5.3.2 and to create a true, hopefully homozygous, CDK12^{as} cell line. Simultaneously, recognizing that the creation of such a cell line will be very challenging, we will also attempt to create a CDK12 knockout cell line (or a shRNA depleted CDK12 cell line) rescued with plasmid expressed CDK12^{as}/CyclinK. As previously touched upon, a CDK12^{as} cell line will allow for the differentiation between catalytic and non-catalytic effects and for the temporal control to parse the immediate and long-term (secondary) consequence of CDK12's inhibition.

Concurrently we are also interested in the verification and further investigation of the CDK12 associated proteins that we identified by mass spectrometry. In addition to further developing our understanding of CDK12 function, we hope to use these associations to further explore the functions of hCDK13. The similarities between the kinase domains of hCDK12 and hCDK13, coupled with the *in vitro* data suggest that hCDK13 is capable of *in vivo* CTD kinase activity. Thus the lack of observed *in vivo*

activity is probably due to the targeting of the hCDK13 kinase to other substrates, splicing factors being the most likely candidates based on current evidence in the literature (Berro et al., 2008; Even et al., 2006; Ko et al., 2001). We expect that the targeting of hCDK12 and 13 is mediated through their divergent N and C terminal extensions, which probably associate with different sets of factors. In order to gain more insight into the *in vivo* activity of both kinases, we will attempt to purify endogenous hCDK13 to identify its associated complexes and proteins. We anticipate that the differences between hCDK12 and hCDK13 associated proteins will be informative.

Finally, in the mid-term, we recognize that a more indepth understanding of the functions of both CDK12 and CDK13 will require the leverage of new, genome wide, deep sequencing techniques that interrogate nascent RNA, such as PRO-Seq. We hope that through the development of inhibitable cell lines and through the deployment of these genome wide techniques, we will be able to contribute to the rapidly growing body of knowledge regarding CDK12, the CTD of RNAPII, and transcription.

Appendix A

Table 4: dCDK12 associate proteins identified by mass spectrometry; proteins present in control immunoprecipitations have been removed from the list with the exception of SC35 (in red, 1 peptide in control).

Identified Proteins	Accession	CDK12 IP: # of peptides identified
brahma protein [Drosophila melanogaster]	gi|157012	56
moira, isoform A [Drosophila melanogaster]	gi|17737997	36
osa, isoform A [Drosophila melanogaster]	gi|24647755	29
FI03643p [Drosophila melanogaster]	gi|218505869	27
brahma associated protein 60kD [Drosophila melanogaster]	gi|24641689	24
brahma associated protein 170kD [Drosophila melanogaster]	gi|19921692	24
Microtubule-associated protein 60 [Drosophila melanogaster]	gi|17136860	22
Brahma-associated protein 111kD [Drosophila melanogaster]	gi|13591766	21
GE14195 [Drosophila yakuba]	gi|195488752	18
scully [Drosophila melanogaster]	gi|17737361	17
Calcium/calmodulin-dependent protein kinase II, isoform B [Drosophila melanogaster]	gi|24638772	17
SF2 [Drosophila melanogaster]	gi|21358097	16
CG7154 [Drosophila melanogaster]	gi|19920888	16
GG12837 [Drosophila erecta]	gi|194898751	12
d4, isoform A [Drosophila melanogaster]	gi|19921648	11
phosphoglycerate mutase 5 [Drosophila melanogaster]	gi|18543249	6
LD10526p [Drosophila melanogaster]	gi|21391996	6
Prp19 [Drosophila melanogaster]	gi|17647459	6
CG9776, isoform A [Drosophila melanogaster]	gi|21356019	6
stonewall [Drosophila melanogaster]	gi|24664166	6
hook-like, isoform A [Drosophila melanogaster]	gi|20129613	6
GM11367 [Drosophila sechellia]	gi|195350482	6
Rrp40 [Drosophila melanogaster]	gi|24580923	5
eIF4AIII [Drosophila melanogaster]]	gi|125775187	5
CG7597 [Drosophila melanogaster]	gi|17862948	5
GM01970p [Drosophila melanogaster]	gi|16183595	5
cyclin K, isoform A [Drosophila melanogaster]	gi|195354127	4
CG1621 [Drosophila melanogaster]	gi|21687080	4
fruitless type-A [Drosophila melanogaster]	gi|11066442	4
GE15779 [Drosophila yakuba]	gi|195480143	4

SC35, isoform B [Drosophila melanogaster]	gi|21358099	3
modifier of mdg4, isoform E [Drosophila melanogaster]	gi|24648734	3
CG6459 [Drosophila melanogaster]	gi|116806184	3
protein on ecdysone puffs, isoform B [Drosophila melanogaster]	gi|17864514	3
GM23813 [Drosophila sechellia]	gi|195330384	3
CG6654 [Drosophila melanogaster]	gi|21357595	3
CG6617 [Drosophila melanogaster]	gi|24643081	3
SRm160 [Drosophila melanogaster]	gi|24663664	3
CG11266, isoform B [Drosophila melanogaster]	gi|19920866	3
heterochromatin protein 5, isoform B [Drosophila melanogaster]	gi|24641302	2
CG2926 [Drosophila melanogaster]	gi|24644293	2
GG15416 [Drosophila erecta]	gi|194868191	2
putative exoribonuclease DIS3 [Drosophila melanogaster]	gi|13446610	2
GG20559 [Drosophila erecta]	gi|194882669	2
GA21150 [Drosophila pseudoobscura pseudoobscura]	gi|125811459	2
Rrp6 [Drosophila melanogaster]	gi|161078302	2
suppressor of variegation 205, isoform A [Drosophila melanogaster]	gi|17136528	2
RE67757p [Drosophila melanogaster]	gi|17946442	2
small ribonucleoprotein particle protein SmG, isoform A [Drosophila melanogaster]	gi|18860007	2
GM15529 [Drosophila sechellia]	gi|195347132	2
CG14711 [Drosophila melanogaster]	gi|24646011	2
pinin, isoform B [Drosophila melanogaster]	gi|161078140	2
xl6 [Drosophila melanogaster]	gi|24582360	2
EG:133E12.4 [Drosophila melanogaster]	gi|2661513	2
Mtr3 [Drosophila melanogaster]	gi|24667071	1
RNA binding protein [Drosophila melanogaster]	gi|158224	1
Rrp45 [Drosophila melanogaster]	gi|24642649	1
toothrin [Drosophila melanogaster]	gi|18859713	1
GG23563 [Drosophila erecta]	gi|194862742	1
diskette [Drosophila melanogaster]	gi|17647341	1
lark, isoform A [Drosophila melanogaster]	gi|17647581	1
Chain A, Structural Analysis Of A Cytoplasmic Dynein Light Chain-Intermediate Chain Complex	gi|149243127	1
actin 87E, isoform A [Drosophila melanogaster]	gi|17137090	1
microtubule dependent motor protein [Drosophila melanogaster]	gi|110808371	1
GA19319 [Drosophila pseudoobscura pseudoobscura]	gi|125777721	1
GA21399 [Drosophila pseudoobscura pseudoobscura]	gi|125809258	1
GA21317 [Drosophila pseudoobscura pseudoobscura]	gi|125809441	1
GA20724 [Drosophila pseudoobscura pseudoobscura]	gi|125977468	1

GA19471 [Drosophila pseudoobscura pseudoobscura]	gi|125978158	1
SD04057p [Drosophila melanogaster]	gi|15292493	1
IP18396p [Drosophila melanogaster]	gi|159884119	1
ribosomal protein S5a [Drosophila silvestris]	gi|164430966	1
SD07884p [Drosophila melanogaster]	gi|16769882	1
CKII-alpha [Drosophila yakuba]	gi|17016246	1
GF21187 [Drosophila ananassae]	gi|194763266	1
GD15571 [Drosophila simulans]	gi|195567753	1
GH16479p [Drosophila melanogaster]	gi|21428924	1
LP02069p [Drosophila melanogaster]	gi|21430330	1
SD22208p [Drosophila melanogaster]	gi|21430870	1
CG15784 [Drosophila melanogaster]	gi|24639869	1
CG8928 [Drosophila melanogaster]	gi|24642361	1

Appendix B

Table 5: hCDK12 associated proteins identified by mass spectrometry; proteins present in control immunoprecipitations have been removed from the list.

Identified Protein	UniProt ID	CDK12 IP: # of peptides identified
Cell division protein kinase 12 OS=Homo sapiens GN=CDK12 PE=1 SV=2	CDK12_HUMAN	95
Serine/arginine repetitive matrix protein 2 OS=Homo sapiens GN=SRRM2 PE=1 SV=2	SRRM2_HUMAN	94
Pre-mRNA cleavage complex 2 protein Pcf11 OS=Homo sapiens GN=PCF11 PE=1 SV=3	PCF11_HUMAN	77
ELKS/Rab6-interacting/CAST family member 1 OS=Homo sapiens GN=ERC1 PE=1 SV=1	RB6I2_HUMAN	61
Protein phosphatase 1 regulatory subunit 12A OS=Homo sapiens GN=PPP1R12A PE=1 SV=1	MYPT1_HUMAN	56
Protein BAT2-like 1 OS=Homo sapiens GN=BAT2L1 PE=1 SV=2	BA2L1_HUMAN	55
Transcription elongation factor B polypeptide 3 OS=Homo sapiens GN=TCEB3 PE=1 SV=2	ELOA1_HUMAN	51
DNA excision repair protein ERCC-6 OS=Homo sapiens GN=ERCC6 PE=1 SV=1	ERCC6_HUMAN	49
Antigen KI-67 OS=Homo sapiens GN=MKI67 PE=1 SV=2	KI67_HUMAN	41
Structural maintenance of chromosomes protein 4 OS=Homo sapiens GN=SMC4 PE=1 SV=2	SMC4_HUMAN	39
Transcriptional regulator ATRX OS=Homo sapiens GN=ATRX PE=1 SV=5	ATRX_HUMAN	38
Polyadenylate-binding protein 1 OS=Homo sapiens GN=PABPC1 PE=1 SV=2	PABP1_HUMAN	38
DNA repair protein RAD50 OS=Homo sapiens GN=RAD50 PE=1 SV=1	RAD50_HUMAN	34
Dedicator of cytokinesis protein 5 OS=Homo sapiens GN=DOCK5 PE=1 SV=3	DOCK5_HUMAN	34
Structural maintenance of chromosomes protein 2 OS=Homo sapiens GN=SMC2 PE=1 SV=2	SMC2_HUMAN	33
Protein VPRBP OS=Homo sapiens GN=VPRBP PE=1 SV=3	VPRBP_HUMAN	33
Apoptotic chromatin condensation inducer in the nucleus OS=Homo sapiens GN=ACIN1 PE=1 SV=2	ACINU_HUMAN	32
U5 small nuclear ribonucleoprotein 200 kDa helicase OS=Homo sapiens GN=SNRNP200 PE=1 SV=2	U520_HUMAN	31
DNA-directed RNA polymerase II subunit RPB1 OS=Homo sapiens GN=POLR2A PE=1 SV=2	RPB1_HUMAN	31
DNA damage-binding protein 1 OS=Homo sapiens GN=DDB1 PE=1 SV=1	DDB1_HUMAN	31
RNA exonuclease 1 homolog OS=Homo sapiens GN=REXO1 PE=1 SV=3	REXO1_HUMAN	29
LIM domain only protein 7 OS=Homo sapiens GN=LMO7 PE=1 SV=3	LMO7_HUMAN	29
Large proline-rich protein BAT2 OS=Homo sapiens GN=BAT2 PE=1 SV=3	BAT2_HUMAN	29
Integrator complex subunit 1 OS=Homo sapiens GN=INTS1 PE=1 SV=2	INT1_HUMAN	29
Protein phosphatase 1 regulatory subunit 12C OS=Homo sapiens GN=PPP1R12C PE=1 SV=1	PP12C_HUMAN	29
BTB/POZ domain-containing protein KCTD3 OS=Homo sapiens GN=KCTD3 PE=1 SV=2	KCTD3_HUMAN	29
Kinesin-like protein KIF23 OS=Homo sapiens GN=KIF23 PE=1 SV=3	KIF23_HUMAN	27
Ankycorbin OS=Homo sapiens GN=RAI14 PE=1 SV=2	RAI14_HUMAN	27
Sorbin and SH3 domain-containing protein 2 OS=Homo sapiens GN=SORBS2 PE=1 SV=3	SRBS2_HUMAN	27
Zinc finger CCCH domain-containing protein 14 OS=Homo sapiens GN=ZC3H14 PE=1 SV=1	ZC3HE_HUMAN	26

Transformation/transcription domain-associated protein OS=Homo sapiens GN=TRRAP PE=1 SV=3	TRRAP_HUMAN	25
Probable E3 ubiquitin-protein ligase MYCBP2 OS=Homo sapiens GN=MYCBP2 PE=1 SV=3	MYCB2_HUMAN	25
Zinc finger CCHC domain-containing protein 8 OS=Homo sapiens GN=ZCCHC8 PE=1 SV=2	ZCHC8_HUMAN	24
Polyribonucleotide 5'-hydroxyl-kinase Clp1 OS=Homo sapiens GN=CLP1 PE=1 SV=1	CLP1_HUMAN	24
Serine/arginine-rich splicing factor 6 OS=Homo sapiens GN=SRSF6 PE=1 SV=2	SRSF6_HUMAN	23
Cyclin-K OS=Homo sapiens GN=CCNK PE=1 SV=2	CCNK_HUMAN	23
Formin-binding protein 1-like OS=Homo sapiens GN=FNBP1L PE=1 SV=2	FBP1L_HUMAN	23
Bcl-2-associated transcription factor 1 OS=Homo sapiens GN=BCLAF1 PE=1 SV=2	BCLF1_HUMAN	22
Tripartite motif-containing protein 3 OS=Homo sapiens GN=TRIM3 PE=1 SV=2	TRIM3_HUMAN	22
Zinc finger and BTB domain-containing protein 11 OS=Homo sapiens GN=ZBTB11 PE=1 SV=2	ZBT11_HUMAN	22
E1A-binding protein p400 OS=Homo sapiens GN=EP400 PE=1 SV=3	EP400_HUMAN	21
TRIO and F-actin-binding protein OS=Homo sapiens GN=TRIOBP PE=1 SV=3	TARA_HUMAN	21
Serine/arginine repetitive matrix protein 1 OS=Homo sapiens GN=SRRM1 PE=1 SV=2	SRRM1_HUMAN	21
Pre-mRNA-processing-splicing factor 8 OS=Homo sapiens GN=PRPF8 PE=1 SV=2	PRP8_HUMAN	20
Myosin phosphatase Rho-interacting protein OS=Homo sapiens GN=MPRIP PE=1 SV=3	MPRIP_HUMAN	20
Zinc finger and BTB domain-containing protein 1 OS=Homo sapiens GN=ZBTB1 PE=1 SV=3	ZBTB1_HUMAN	20
Thyroid hormone receptor-associated protein 3 OS=Homo sapiens GN=THRAP3 PE=1 SV=2	TR150_HUMAN	19
116 kDa U5 small nuclear ribonucleoprotein component OS=Homo sapiens GN=EFTUD2 PE=1 SV=1	U5S1_HUMAN	19
Zinc finger protein 318 OS=Homo sapiens GN=ZNF318 PE=1 SV=2	ZN318_HUMAN	19
Uveal autoantigen with coiled-coil domains and ankyrin repeats OS=Homo sapiens GN=UACA PE=1 SV=2	UACA_HUMAN	19
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Homo sapiens GN=DHX15 PE=1 SV=2	DHX15_HUMAN	18
Eukaryotic translation initiation factor 4 gamma 1 OS=Homo sapiens GN=EIF4G1 PE=1 SV=4	IF4G1_HUMAN	18
Superkiller viralicidic activity 2-like 2 OS=Homo sapiens GN=SKIV2L2 PE=1 SV=3	SK2L2_HUMAN	18
Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4	IF4A3_HUMAN	18
PRKC apoptosis WT1 regulator protein OS=Homo sapiens GN=PAWR PE=1 SV=1	PAWR_HUMAN	18
Protein Wiz OS=Homo sapiens GN=WIZ PE=1 SV=2	WIZ_HUMAN	18
Nuclease-sensitive element-binding protein 1 OS=Homo sapiens GN=YBX1 PE=1 SV=3	YBOX1_HUMAN	17
La-related protein 1 OS=Homo sapiens GN=LARP1 PE=1 SV=2	LARP1_HUMAN	17
Zinc finger protein ubi-d4 OS=Homo sapiens GN=DPF2 PE=1 SV=2	REQU_HUMAN	17
RNA-binding protein 27 OS=Homo sapiens GN=RBM27 PE=1 SV=2	RBM27_HUMAN	16
Polyadenylate-binding protein 4 OS=Homo sapiens GN=PABPC4 PE=1 SV=1	PABP4_HUMAN	16
Serine/threonine-protein phosphatase PP1-beta catalytic subunit OS=Homo sapiens GN=PPP1CB PE=1 SV=3	PP1B_HUMAN	16
Choline-phosphate cytidylyltransferase A OS=Homo sapiens GN=PCYT1A PE=1 SV=2	PCY1A_HUMAN	16
A-kinase anchor protein 12 OS=Homo sapiens GN=AKAP12 PE=1 SV=3	AKA12_HUMAN	16
Splicing factor 3B subunit 3 OS=Homo sapiens GN=SF3B3 PE=1 SV=4	SF3B3_HUMAN	15
Ataxin-2-like protein OS=Homo sapiens GN=ATXN2L PE=1 SV=2	ATX2L_HUMAN	15

40S ribosomal protein S3a OS=Homo sapiens GN=RPS3A PE=1 SV=2	RS3A_HUMAN	15
Matrin-3 OS=Homo sapiens GN=MATR3 PE=1 SV=2	MATR3_HUMAN	15
E3 ubiquitin-protein ligase RNF168 OS=Homo sapiens GN=RNF168 PE=1 SV=1	RN168_HUMAN	15
PiggyBac transposable element-derived protein 3 OS=Homo sapiens GN=PGBD3 PE=2 SV=3	PGBD3_HUMAN	15
Engulfment and cell motility protein 2 OS=Homo sapiens GN=ELMO2 PE=1 SV=2	ELMO2_HUMAN	14
5'-3' exoribonuclease 2 OS=Homo sapiens GN=XRN2 PE=1 SV=1	XRN2_HUMAN	14
Pinin OS=Homo sapiens GN=PNN PE=1 SV=4	PININ_HUMAN	14
G patch domain-containing protein 8 OS=Homo sapiens GN=GPATCH8 PE=1 SV=2	GPTC8_HUMAN	14
SH3KBP1-binding protein 1 OS=Homo sapiens GN=SHKBP1 PE=1 SV=2	SHKB1_HUMAN	14
Histone-lysine N-methyltransferase, H3 lysine-9 specific 5 OS=Homo sapiens GN=EHMT1 PE=1 SV=3	EHMT1_HUMAN	13
Protein arginine N-methyltransferase 5 OS=Homo sapiens GN=PRMT5 PE=1 SV=4	ANM5_HUMAN	13
Nibrin OS=Homo sapiens GN=NBN PE=1 SV=1	NBN_HUMAN	13
Splicing factor 3B subunit 1 OS=Homo sapiens GN=SF3B1 PE=1 SV=3	SF3B1_HUMAN	13
Supervillin OS=Homo sapiens GN=SVIL PE=1 SV=2	SVIL_HUMAN	13
Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=3	HNRPM_HUMAN	13
Tyrosine-protein kinase SgK223 OS=Homo sapiens GN=SGK223 PE=1 SV=2	SG223_HUMAN	13
Microprocessor complex subunit DGCR8 OS=Homo sapiens GN=DGCR8 PE=1 SV=1	DGCR8_HUMAN	13
Histone-lysine N-methyltransferase, H3 lysine-9 specific 3 OS=Homo sapiens GN=EHMT2 PE=1 SV=3	EHMT2_HUMAN	12
Protein capicua homolog OS=Homo sapiens GN=CIC PE=1 SV=2	CIC_HUMAN	12
Sorbin and SH3 domain-containing protein 1 OS=Homo sapiens GN=SORBS1 PE=1 SV=2	SRBS1_HUMAN	12
Transcription elongation factor B polypeptide 2 OS=Homo sapiens GN=TCEB2 PE=1 SV=1	ELOB_HUMAN	12
Nuclear receptor corepressor 1 OS=Homo sapiens GN=NCOR1 PE=1 SV=2	NCOR1_HUMAN	12
Double-strand break repair protein MRE11A OS=Homo sapiens GN=MRE11A PE=1 SV=3	MRE11_HUMAN	11
Exosome component 10 OS=Homo sapiens GN=EXOSC10 PE=1 SV=2	EXOSX_HUMAN	11
SNW domain-containing protein 1 OS=Homo sapiens GN=SNW1 PE=1 SV=1	SNW1_HUMAN	11
RAB6-interacting golgin OS=Homo sapiens GN=GORAB PE=1 SV=1	GORAB_HUMAN	11
Round spermatid basic protein 1 OS=Homo sapiens GN=RSBN1 PE=2 SV=2	RSBN1_HUMAN	11
MAP7 domain-containing protein 3 OS=Homo sapiens GN=MAP7D3 PE=1 SV=2	MA7D3_HUMAN	11
Rac GTPase-activating protein 1 OS=Homo sapiens GN=RACGAP1 PE=1 SV=1	RGAP1_HUMAN	11
Cell cycle regulator Mat89Bb homolog OS=Homo sapiens GN=C12orf11 PE=1 SV=2	M89BB_HUMAN	11
Zinc finger protein 185 OS=Homo sapiens GN=ZNF185 PE=1 SV=3	ZN185_HUMAN	10
UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit OS=Homo sapiens GN=OGT PE=1 SV=3	OGT1_HUMAN	10
Serine/arginine-rich splicing factor 3 OS=Homo sapiens GN=SRSF3 PE=1 SV=1	SRSF3_HUMAN	10
Integrator complex subunit 12 OS=Homo sapiens GN=INTS12 PE=1 SV=1	INT12_HUMAN	10
Serine/threonine-protein phosphatase 1 regulatory subunit 10 OS=Homo sapiens GN=PPP1R10 PE=1 SV=1	PP1RA_HUMAN	10
Cell division cycle 5-like protein OS=Homo sapiens GN=CDC5L PE=1 SV=2	CDC5L_HUMAN	10
CDKN2A-interacting protein OS=Homo sapiens GN=CDKN2AIP PE=1 SV=2	CARF_HUMAN	10

Transformer-2 protein homolog beta OS=Homo sapiens GN=TRA2B PE=1 SV=1	TRA2B_HUMAN	10
Splicing factor 3B subunit 2 OS=Homo sapiens GN=SF3B2 PE=1 SV=2	SF3B2_HUMAN	10
Nuclear fragile X mental retardation-interacting protein 2 OS=Homo sapiens GN=NUFIP2 PE=1 SV=1	NUFP2_HUMAN	9
Tuftelin-interacting protein 11 OS=Homo sapiens GN=TFIP11 PE=1 SV=1	TFP11_HUMAN	9
40S ribosomal protein S8 OS=Homo sapiens GN=RPS8 PE=1 SV=2	RS8_HUMAN	9
Splicing factor 3A subunit 1 OS=Homo sapiens GN=SF3A1 PE=1 SV=1	SF3A1_HUMAN	9
Coiled-coil domain-containing protein 6 OS=Homo sapiens GN=CCDC6 PE=1 SV=2	CCDC6_HUMAN	9
Condensin-2 complex subunit H2 OS=Homo sapiens GN=NCAPH2 PE=1 SV=1	CNDH2_HUMAN	9
Round spermatid basic protein 1-like protein OS=Homo sapiens GN=RSBN1L PE=1 SV=2	RSBNL_HUMAN	9
Zinc finger CCCH domain-containing protein 18 OS=Homo sapiens GN=ZC3H18 PE=1 SV=2	ZCH18_HUMAN	8
Chromobox protein homolog 8 OS=Homo sapiens GN=CBX8 PE=1 SV=3	CBX8_HUMAN	8
SAGA-associated factor 29 homolog OS=Homo sapiens GN=CCDC101 PE=1 SV=1	SGF29_HUMAN	8
40S ribosomal protein S3 OS=Homo sapiens GN=RPS3 PE=1 SV=2	RS3_HUMAN	8
Serine/arginine-rich splicing factor 1 OS=Homo sapiens GN=SRSF1 PE=1 SV=2	SRSF1_HUMAN	8
Nuclear cap-binding protein subunit 1 OS=Homo sapiens GN=NCBP1 PE=1 SV=1	NCBP1_HUMAN	8
Ensconsin OS=Homo sapiens GN=MAP7 PE=1 SV=1	MAP7_HUMAN	8
Serine/arginine-rich splicing factor 7 OS=Homo sapiens GN=SRSF7 PE=1 SV=1	SRSF7_HUMAN	8
E3 ubiquitin-protein ligase RBBP6 OS=Homo sapiens GN=RBBP6 PE=1 SV=1	RBBP6_HUMAN	7
Pre-mRNA-processing factor 19 OS=Homo sapiens GN=PRPF19 PE=1 SV=1	PRP19_HUMAN	7
E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=1 SV=2	UBR5_HUMAN	7
Mitogen-activated protein kinase kinase kinase 7 OS=Homo sapiens GN=MAP3K7 PE=1 SV=1	M3K7_HUMAN	7
Serine/arginine-rich splicing factor 10 OS=Homo sapiens GN=SRSF10 PE=1 SV=1	SRS10_HUMAN	7
DNA-directed RNA polymerase II subunit RPB2 OS=Homo sapiens GN=POLR2B PE=1 SV=1	RPB2_HUMAN	7
WD repeat-containing protein 82 OS=Homo sapiens GN=WDR82 PE=1 SV=1	WDR82_HUMAN	7
Polyadenylate-binding protein 2 OS=Homo sapiens GN=PABPN1 PE=1 SV=3	PABP2_HUMAN	7
Integrator complex subunit 6 OS=Homo sapiens GN=INTS6 PE=1 SV=1	INT6_HUMAN	7
Putative RNA-binding protein Luc7-like 2 OS=Homo sapiens GN=LUC7L2 PE=1 SV=2	LC7L2_HUMAN	7
GC-rich sequence DNA-binding factor 1 OS=Homo sapiens GN=GCFC1 PE=1 SV=2	GCFC1_HUMAN	7
Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC PE=1 SV=4	HNRPC_HUMAN	7
ERC protein 2 OS=Homo sapiens GN=ERC2 PE=1 SV=3	ERC2_HUMAN	7
Ataxin-1-like OS=Homo sapiens GN=ATXN1L PE=1 SV=1	ATX1L_HUMAN	7
Condensin-2 complex subunit D3 OS=Homo sapiens GN=NCAPD3 PE=1 SV=2	CNDD3_HUMAN	7
40S ribosomal protein S7 OS=Homo sapiens GN=RPS7 PE=1 SV=1	RS7_HUMAN	7
Crooked neck-like protein 1 OS=Homo sapiens GN=CRNKL1 PE=1 SV=4	CRNL1_HUMAN	7
Zinc finger protein 644 OS=Homo sapiens GN=ZNF644 PE=1 SV=2	ZN644_HUMAN	7
RNA-binding protein 7 OS=Homo sapiens GN=RBM7 PE=1 SV=1	RBM7_HUMAN	7
A-kinase anchor protein 17A OS=Homo sapiens GN=AKAP17A PE=1 SV=2	AK17A_HUMAN	7

Deoxynucleotidyltransferase terminal-interacting protein 1 OS=Homo sapiens GN=DNTTIP1 PE=1 SV=2	TDIF1_HUMAN	6
Pre-mRNA-processing factor 6 OS=Homo sapiens GN=PRPF6 PE=1 SV=1	PRP6_HUMAN	6
40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3	RS18_HUMAN	6
Exosome complex exonuclease RRP42 OS=Homo sapiens GN=EXOSC7 PE=1 SV=2	EXOS7_HUMAN	6
Histone-binding protein RBBP7 OS=Homo sapiens GN=RBBP7 PE=1 SV=1	RBBP7_HUMAN	6
UAP56-interacting factor OS=Homo sapiens GN=FYTTD1 PE=1 SV=3	UIF_HUMAN	6
Heterogeneous nuclear ribonucleoprotein U-like protein 1 OS=Homo sapiens GN=HNRNPUL1 PE=1 SV=2	HNRL1_HUMAN	6
WD40 repeat-containing protein SMU1 OS=Homo sapiens GN=SMU1 PE=1 SV=2	SMU1_HUMAN	6
Integrator complex subunit 9 OS=Homo sapiens GN=INTS9 PE=1 SV=2	INT9_HUMAN	6
Serine/arginine-rich splicing factor 4 OS=Homo sapiens GN=SRSF4 PE=1 SV=2	SRSF4_HUMAN	6
U2-associated protein SR140 OS=Homo sapiens GN=SR140 PE=1 SV=2	SR140_HUMAN	6
La-related protein 7 OS=Homo sapiens GN=LARP7 PE=1 SV=1	LARP7_HUMAN	6
Histone deacetylase complex subunit SAP18 OS=Homo sapiens GN=SAP18 PE=1 SV=1	SAP18_HUMAN	6
Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3	G3P_HUMAN	6
Heterogeneous nuclear ribonucleoprotein F OS=Homo sapiens GN=HNRNPF PE=1 SV=3	HNRPF_HUMAN	6
Ubiquitin-associated protein 2-like OS=Homo sapiens GN=UBAP2L PE=1 SV=2	UBP2L_HUMAN	6
40S ribosomal protein S2 OS=Homo sapiens GN=RPS2 PE=1 SV=2	RS2_HUMAN	6
DNA-binding protein A OS=Homo sapiens GN=CSDA PE=1 SV=4	DBPA_HUMAN	6
40S ribosomal protein S12 OS=Homo sapiens GN=RPS12 PE=1 SV=3	RS12_HUMAN	6
60S ribosomal protein L13 OS=Homo sapiens GN=RPL13 PE=1 SV=4	RL13_HUMAN	6
40S ribosomal protein S9 OS=Homo sapiens GN=RPS9 PE=1 SV=3	RS9_HUMAN	6
40S ribosomal protein S5 OS=Homo sapiens GN=RPS5 PE=1 SV=4	RS5_HUMAN	6
40S ribosomal protein S16 OS=Homo sapiens GN=RPS16 PE=1 SV=2	RS16_HUMAN	6
40S ribosomal protein S14 OS=Homo sapiens GN=RPS14 PE=1 SV=3	RS14_HUMAN	6
Serine/arginine-rich splicing factor 9 OS=Homo sapiens GN=SRSF9 PE=1 SV=1	SRSF9_HUMAN	6
Mitochondrial fission regulator 1 OS=Homo sapiens GN=MTFR1 PE=1 SV=2	MTFR1_HUMAN	5
TGF-beta-activated kinase 1 and MAP3K7-binding protein 1 OS=Homo sapiens GN=TAB1 PE=1 SV=1	TAB1_HUMAN	5
3'-5' exoribonuclease CSL4 homolog OS=Homo sapiens GN=EXOSC1 PE=1 SV=1	EXOS1_HUMAN	5
RNA-binding protein 8A OS=Homo sapiens GN=RBM8A PE=1 SV=1	RBM8A_HUMAN	5
SUN domain-containing protein 2 OS=Homo sapiens GN=SUN2 PE=1 SV=3	SUN2_HUMAN	5
Cell division protein kinase 13 OS=Homo sapiens GN=CDK13 PE=1 SV=2	CDK13_HUMAN	5
Structural maintenance of chromosomes protein 1A OS=Homo sapiens GN=SMC1A PE=1 SV=2	SMC1A_HUMAN	5
Calcium homeostasis endoplasmic reticulum protein OS=Homo sapiens GN=CHERP PE=1 SV=3	CHERP_HUMAN	5
E3 ubiquitin-protein ligase RING2 OS=Homo sapiens GN=RNF2 PE=1 SV=1	RING2_HUMAN	5
Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2	H2AZ_HUMAN	5
40S ribosomal protein S23 OS=Homo sapiens GN=RPS23 PE=1 SV=3	RS23_HUMAN	5
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial OS=Homo sapiens GN=MCCC2 PE=1 SV=1	MCCB_HUMAN	5

E3 SUMO-protein ligase CBX4 OS=Homo sapiens GN=CBX4 PE=1 SV=3	CBX4_HUMAN	5
Nucleophosmin OS=Homo sapiens GN=NPM1 PE=1 SV=2	NPM_HUMAN	5
Heat shock 70 kDa protein 6 OS=Homo sapiens GN=HSPA6 PE=1 SV=2	HSP76_HUMAN	4
40S ribosomal protein S11 OS=Homo sapiens GN=RPS11 PE=1 SV=3	RS11_HUMAN	4
40S ribosomal protein S6 OS=Homo sapiens GN=RPS6 PE=1 SV=1	RS6_HUMAN	4
40S ribosomal protein S25 OS=Homo sapiens GN=RPS25 PE=1 SV=1	RS25_HUMAN	4
Serine/arginine-rich splicing factor 5 OS=Homo sapiens GN=SRSF5 PE=1 SV=1	SRSF5_HUMAN	4
Transformer-2 protein homolog alpha OS=Homo sapiens GN=TRA2A PE=1 SV=1	TRA2A_HUMAN	4
Transcription elongation factor B polypeptide 1 OS=Homo sapiens GN=TCEB1 PE=1 SV=1	ELOC_HUMAN	4
Exosome complex exonuclease RRP43 OS=Homo sapiens GN=EXOSC8 PE=1 SV=1	EXOS8_HUMAN	4
Uncharacterized protein C9orf82 OS=Homo sapiens GN=C9orf82 PE=1 SV=2	CI082_HUMAN	4
Exosome complex exonuclease RRP46 OS=Homo sapiens GN=EXOSC5 PE=1 SV=1	EXOS5_HUMAN	4
Probable ribosome biogenesis protein NEP1 OS=Homo sapiens GN=EMG1 PE=1 SV=4	NEP1_HUMAN	4
Pre-mRNA-splicing factor SYF2 OS=Homo sapiens GN=SYF2 PE=1 SV=1	SYF2_HUMAN	4
Putative tubulin beta-4q chain OS=Homo sapiens GN=TUBB4Q PE=5 SV=1	TBB4Q_HUMAN	4
Exosome complex exonuclease MTR3 OS=Homo sapiens GN=EXOSC6 PE=1 SV=1	EXOS6_HUMAN	4
Pleiotropic regulator 1 OS=Homo sapiens GN=PLRG1 PE=1 SV=1	PLRG1_HUMAN	4
Luc7-like protein 3 OS=Homo sapiens GN=LUC7L3 PE=1 SV=2	LC7L3_HUMAN	4
Nuclear cap-binding protein subunit 2 OS=Homo sapiens GN=NCBP2 PE=1 SV=1	NCBP2_HUMAN	4
Integrator complex subunit 8 OS=Homo sapiens GN=INTS8 PE=1 SV=1	INT8_HUMAN	4
Actin-like protein 6A OS=Homo sapiens GN=ACTL6A PE=1 SV=1	ACL6A_HUMAN	4
Integrator complex subunit 7 OS=Homo sapiens GN=INTS7 PE=1 SV=1	INT7_HUMAN	4
Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=2	ILF2_HUMAN	4
Histone deacetylase 3 OS=Homo sapiens GN=HDAC3 PE=1 SV=2	HDAC3_HUMAN	4
Dual specificity protein kinase CLK3 OS=Homo sapiens GN=CLK3 PE=1 SV=3	CLK3_HUMAN	3
Treslin OS=Homo sapiens GN=TICRR PE=1 SV=2	TICRR_HUMAN	3
Putative RNA-binding protein 15 OS=Homo sapiens GN=RBM15 PE=1 SV=2	RBM15_HUMAN	3
Protein SON OS=Homo sapiens GN=SON PE=1 SV=4	SON_HUMAN	3
Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16 OS=Homo sapiens GN=DHX38 PE=1 SV=2	PRP16_HUMAN	3
Methyl-CpG-binding protein 2 OS=Homo sapiens GN=MECP2 PE=1 SV=1	MECP2_HUMAN	3
Splicing factor U2AF 65 kDa subunit OS=Homo sapiens GN=U2AF2 PE=1 SV=4	U2AF2_HUMAN	3
Heterogeneous nuclear ribonucleoprotein K OS=Homo sapiens GN=HNRNPK PE=1 SV=1	HNRPK_HUMAN	3
40S ribosomal protein S17 OS=Homo sapiens GN=RPS17 PE=1 SV=2	RS17_HUMAN	3
60S ribosomal protein L12 OS=Homo sapiens GN=RPL12 PE=1 SV=1	RL12_HUMAN	3
40S ribosomal protein S19 OS=Homo sapiens GN=RPS19 PE=1 SV=2	RS19_HUMAN	3
Small nuclear ribonucleoprotein Sm D1 OS=Homo sapiens GN=SNRPD1 PE=1 SV=1	SMD1_HUMAN	3
RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=1 SV=1	EWS_HUMAN	3
Akirin-2 OS=Homo sapiens GN=AKIRIN2 PE=1 SV=2	AKIR2_HUMAN	3

Protein DGCR6L OS=Homo sapiens GN=DGCR6L PE=1 SV=2	DGC6L_HUMAN	3
60S ribosomal protein L27a OS=Homo sapiens GN=RPL27A PE=1 SV=2	RL27A_HUMAN	3
40S ribosomal protein S15a OS=Homo sapiens GN=RPS15A PE=1 SV=2	RS15A_HUMAN	3
Exosome complex exonuclease RRP45 OS=Homo sapiens GN=EXOSC9 PE=1 SV=3	EXOS9_HUMAN	3
Ribosomal L1 domain-containing protein 1 OS=Homo sapiens GN=RSL1D1 PE=1 SV=3	RL1D1_HUMAN	3
Serine/Arginine-related protein 53 OS=Homo sapiens GN=RSRC1 PE=1 SV=1	RSRC1_HUMAN	3
40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=2	RS13_HUMAN	3
TOX high mobility group box family member 4 OS=Homo sapiens GN=TOX4 PE=1 SV=1	TOX4_HUMAN	3
Exosome complex exonuclease RRP4 OS=Homo sapiens GN=EXOSC2 PE=1 SV=2	EXOS2_HUMAN	3
Protein mago nashi homolog OS=Homo sapiens GN=MAGOH PE=1 SV=1	MGN_HUMAN	3
DNA-directed RNA polymerase II subunit RPB3 OS=Homo sapiens GN=POLR2C PE=1 SV=2	RPB3_HUMAN	3
DNA-directed RNA polymerases I, II, and III subunit RPABC1 OS=Homo sapiens GN=POLR2E PE=1 SV=4	RPAB1_HUMAN	3
Ribonuclease P protein subunit p30 OS=Homo sapiens GN=RPP30 PE=1 SV=1	RPP30_HUMAN	3
SWI/SNF complex subunit SMARCC1 OS=Homo sapiens GN=SMARCC1 PE=1 SV=3	SMRC1_HUMAN	3
ATP-dependent RNA helicase DDX3X OS=Homo sapiens GN=DDX3X PE=1 SV=3	DDX3X_HUMAN	3
Histone-lysine N-methyltransferase SETD2 OS=Homo sapiens GN=SETD2 PE=1 SV=3	SETD2_HUMAN	3
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	CSK21_HUMAN	3
MAP3K12-binding inhibitory protein 1 OS=Homo sapiens GN=MBIP PE=1 SV=2	MBIP1_HUMAN	3
F-box-like/WD repeat-containing protein TBL1XR1 OS=Homo sapiens GN=TBL1XR1 PE=1 SV=1	TBL1R_HUMAN	3
40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1	RS20_HUMAN	3
Dynein light chain 1, cytoplasmic OS=Homo sapiens GN=DYNLL1 PE=1 SV=1	DYL1_HUMAN	3
Enhancer of rudimentary homolog OS=Homo sapiens GN=ERH PE=1 SV=1	ERH_HUMAN	3
Pre-mRNA-splicing factor SPF27 OS=Homo sapiens GN=BCAS2 PE=1 SV=1	SPF27_HUMAN	3
Histone H2B type F-S OS=Homo sapiens GN=H2BFS PE=1 SV=2	H2BFS_HUMAN	3
UPF0468 protein C16orf80 OS=Homo sapiens GN=C16orf80 PE=1 SV=1	CP080_HUMAN	3
U5 small nuclear ribonucleoprotein 40 kDa protein OS=Homo sapiens GN=SNRNP40 PE=1 SV=1	SNR40_HUMAN	3
PHD finger-like domain-containing protein 5A OS=Homo sapiens GN=PHF5A PE=1 SV=1	PHF5A_HUMAN	3
Methylosome protein 50 OS=Homo sapiens GN=WDR77 PE=1 SV=1	MEP50_HUMAN	3
Polyadenylate-binding protein 3 OS=Homo sapiens GN=PABPC3 PE=1 SV=2	PABP3_HUMAN	3
Bromodomain-containing protein 8 OS=Homo sapiens GN=BRD8 PE=1 SV=2	BRD8_HUMAN	3
Protein CASP OS=Homo sapiens GN=CUX1 PE=1 SV=2	CASP_HUMAN	3
Splicing factor 3B subunit 4 OS=Homo sapiens GN=SF3B4 PE=1 SV=1	SF3B4_HUMAN	3
Histone H1.3 OS=Homo sapiens GN=HIST1H1D PE=1 SV=2	H13_HUMAN	2
Probable ATP-dependent RNA helicase DDX41 OS=Homo sapiens GN=DDX41 PE=1 SV=2	DDX41_HUMAN	2
60S ribosomal protein L18 OS=Homo sapiens GN=RPL18 PE=1 SV=2	RL18_HUMAN	2
Pre-mRNA-processing factor 17 OS=Homo sapiens GN=CDC40 PE=1 SV=1	PRP17_HUMAN	2
Casein kinase II subunit beta OS=Homo sapiens GN=CSNK2B PE=1 SV=1	CSK2B_HUMAN	2

Phosphofurin acidic cluster sorting protein 1 OS=Homo sapiens GN=PACS1 PE=1 SV=2	PACS1_HUMAN	2
Choline-phosphate cytidylyltransferase B OS=Homo sapiens GN=PCYT1B PE=1 SV=1	PCY1B_HUMAN	2
Insulin-like growth factor 2 mRNA-binding protein 3 OS=Homo sapiens GN=IGF2BP3 PE=1 SV=2	IF2B3_HUMAN	2
Cleavage and polyadenylation specificity factor subunit 1 OS=Homo sapiens GN=CPSF1 PE=1 SV=2	CPSF1_HUMAN	2
Vacuolar protein sorting-associated protein 72 homolog OS=Homo sapiens GN=VPS72 PE=1 SV=1	VPS72_HUMAN	2
Small nuclear ribonucleoprotein Sm D3 OS=Homo sapiens GN=SNRPD3 PE=1 SV=1	SMD3_HUMAN	2
Serine/arginine-rich splicing factor 8 OS=Homo sapiens GN=SRSF8 PE=1 SV=1	SRSF8_HUMAN	2
60S ribosomal protein L29 OS=Homo sapiens GN=RPL29 PE=1 SV=2	RL29_HUMAN	2
F-box/SPRY domain-containing protein 1 OS=Homo sapiens GN=FBXO45 PE=1 SV=1	FBSP1_HUMAN	2
60S ribosomal protein L22 OS=Homo sapiens GN=RPL22 PE=1 SV=2	RL22_HUMAN	2
40S ribosomal protein S10 OS=Homo sapiens GN=RPS10 PE=1 SV=1	RS10_HUMAN	2
Serine/threonine-protein phosphatase PP1-alpha catalytic subunit OS=Homo sapiens GN=PPP1CA PE=1 SV=1	PP1A_HUMAN	2
40S ribosomal protein S26 OS=Homo sapiens GN=RPS26 PE=1 SV=3	RS26_HUMAN	2
60S ribosomal protein L30 OS=Homo sapiens GN=RPL30 PE=1 SV=2	RL30_HUMAN	2
60S ribosomal protein L31 OS=Homo sapiens GN=RPL31 PE=1 SV=1	RL31_HUMAN	2
Protein DGCR6 OS=Homo sapiens GN=DGCR6 PE=1 SV=3	DGCR6_HUMAN	2
Reticulocalbin-1 OS=Homo sapiens GN=RCN1 PE=1 SV=1	RCN1_HUMAN	2
DET1- and DDB1-associated protein 1 OS=Homo sapiens GN=DDA1 PE=1 SV=1	DDA1_HUMAN	2
Pre-mRNA branch site protein p14 OS=Homo sapiens GN=SF3B14 PE=1 SV=1	PM14_HUMAN	2
Putative RNA-binding protein Luc7-like 1 OS=Homo sapiens GN=LUC7L PE=1 SV=1	LUC7L_HUMAN	2
RNA-binding protein 25 OS=Homo sapiens GN=RBM25 PE=1 SV=3	RBM25_HUMAN	2
Cell division protein kinase 9 OS=Homo sapiens GN=CDK9 PE=1 SV=3	CDK9_HUMAN	2
Pre-mRNA-splicing factor 38A OS=Homo sapiens GN=PRPF38A PE=1 SV=1	PR38A_HUMAN	2
60S ribosomal protein L9 OS=Homo sapiens GN=RPL9 PE=1 SV=1	RL9_HUMAN	2
Protein LSM12 homolog OS=Homo sapiens GN=LSM12 PE=1 SV=2	LSM12_HUMAN	2
Male-specific lethal 1 homolog OS=Homo sapiens GN=MSL1 PE=1 SV=2	MSL1_HUMAN	2
Eukaryotic initiation factor 4A-I OS=Homo sapiens GN=EIF4A1 PE=1 SV=1	IF4A1_HUMAN	2
60S ribosomal protein L10a OS=Homo sapiens GN=RPL10A PE=1 SV=2	RL10A_HUMAN	2
Nucleolar protein 9 OS=Homo sapiens GN=NOL9 PE=1 SV=1	NOL9_HUMAN	2
Pre-mRNA 3'-end-processing factor FIP1 OS=Homo sapiens GN=FIP1L1 PE=1 SV=1	FIP1_HUMAN	2
Glutamine and serine-rich protein 1 OS=Homo sapiens GN=QSER1 PE=1 SV=3	QSER1_HUMAN	2
NF-kappa-B-repressing factor OS=Homo sapiens GN=NKRF PE=1 SV=2	NKRF_HUMAN	2
Replication factor C subunit 1 OS=Homo sapiens GN=RFC1 PE=1 SV=4	RFC1_HUMAN	2
60S ribosomal protein L18a OS=Homo sapiens GN=RPL18A PE=1 SV=2	RL18A_HUMAN	2
RNA-binding protein with serine-rich domain 1 OS=Homo sapiens GN=RNPS1 PE=1 SV=1	RNPS1_HUMAN	2
60S acidic ribosomal protein P1 OS=Homo sapiens GN=RPLP1 PE=1 SV=1	RLA1_HUMAN	2
60S ribosomal protein L35a OS=Homo sapiens GN=RPL35A PE=1 SV=2	RL35A_HUMAN	2

RNA-binding protein FUS OS=Homo sapiens GN=FUS PE=1 SV=1	FUS_HUMAN	2
Serine/threonine-protein phosphatase PP1-gamma catalytic subunit OS=Homo sapiens GN=PPP1CC PE=1 SV=1	PP1G_HUMAN	2
40S ribosomal protein S27 OS=Homo sapiens GN=RPS27 PE=1 SV=3	RS27_HUMAN	2
60S ribosomal protein L21 OS=Homo sapiens GN=RPL21 PE=1 SV=2	RL21_HUMAN	2
60S ribosomal protein L27 OS=Homo sapiens GN=RPL27 PE=1 SV=2	RL27_HUMAN	2
S-phase kinase-associated protein 1 OS=Homo sapiens GN=SKP1 PE=1 SV=2	SKP1_HUMAN	2
Large neutral amino acids transporter small subunit 1 OS=Homo sapiens GN=SLC7A5 PE=1 SV=2	LAT1_HUMAN	2
Complement component 1 Q subcomponent-binding protein, mitochondrial OS=Homo sapiens GN=C1QBP PE=1 SV=1	C1QBP_HUMAN	2
60S acidic ribosomal protein P0-like OS=Homo sapiens GN=RPLP0P6 PE=5 SV=1	RLA0L_HUMAN	2
F-box-like/WD repeat-containing protein TBL1Y OS=Homo sapiens GN=TBL1Y PE=2 SV=1	TBL1Y_HUMAN	2
Zinc finger protein 576 OS=Homo sapiens GN=ZNF576 PE=1 SV=1	ZN576_HUMAN	2
RNA-binding protein Raly OS=Homo sapiens GN=RALY PE=1 SV=1	RALY_HUMAN	2
Poly(U)-binding-splicing factor PUF60 OS=Homo sapiens GN=PUF60 PE=1 SV=1	PUF60_HUMAN	2
Exosome complex exonuclease RRP41 OS=Homo sapiens GN=EXOSC4 PE=1 SV=3	EXOS4_HUMAN	2
L-lactate dehydrogenase B chain OS=Homo sapiens GN=LDHB PE=1 SV=2	LDHB_HUMAN	2
60S ribosomal protein L8 OS=Homo sapiens GN=RPL8 PE=1 SV=2	RL8_HUMAN	2
Polyadenylate-binding protein 1-like OS=Homo sapiens GN=PABPC1L PE=2 SV=1	PAP1L_HUMAN	2
60S ribosomal protein L17 OS=Homo sapiens GN=RPL17 PE=1 SV=3	RL17_HUMAN	2
26S protease regulatory subunit 4 OS=Homo sapiens GN=PSMC1 PE=1 SV=1	PRS4_HUMAN	2
Integrator complex subunit 5 OS=Homo sapiens GN=INTS5 PE=1 SV=1	INT5_HUMAN	2

References

Ahn, S. H., Keogh, M. C., Buratowski, S., 2009. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J. 28, 205-12.

Ahn, S. H., Kim, M., Buratowski, S., 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell. 13, 67-76.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., Ansari, A. Z., 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell. 34, 387-93.

Allison, L. A., Moyle, M., Shales, M., Ingles, C. J., 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell. 42, 599-610.

Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., Ingles, C. J., 1988. The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. Mol Cell Biol. 8, 321-9.

Allo, M., Buggiano, V., Fededa, J. P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S. A., Klinck, R., Chabot, B., Kornblihtt, A. R., 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol. 16, 717-24.

Armstrong, J. A., Papoulas, O., Daubresse, G., Sperling, A. S., Lis, J. T., Scott, M. P., Tamkun, J. W., 2002. The Drosophila BRM complex facilitates global transcription by RNA polymerase II. EMBO J. 21, 5245-54.

Baillat, D., Hakimi, M. A., Naar, A. M., Shilatifard, A., Cooch, N., Shiekhattar, R., 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 123, 265-76.

Bajrami, I., Frankum, J. R., Konde, A., Miller, R. E., Rehman, F. L., Brough, R., Campbell, J., Sims, D., Rafiq, R., Hooper, S., Chen, L., Kozarewa, I., Assiotis, I., Fenwick, K., Natrajan, R., Lord, C. J., Ashworth, A., 2014. Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. Cancer Res. 74, 287-97.

Barilla, D., Lee, B. A., Proudfoot, N. J., 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 98, 445-50.

Bartholomew, B., Dahmus, M. E., Meares, C. F., 1986. RNA contacts subunits IIo and IIc in HeLa RNA polymerase II transcription complexes. J Biol Chem. 261, 14226-31.

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., Greenleaf, A. L., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24, 2303-16.

Bartkowiak, B., Mackellar, A. L., Greenleaf, A. L., 2011. Updating the CTD Story: From Tail to Epic. Genet Res Int. 2011, 623718.

Baskaran, R., Dahmus, M. E., Wang, J. Y., 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proc Natl Acad Sci U S A. 90, 11167-71.

Batsche, E., Yaniv, M., Muchardt, C., 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nat Struct Mol Biol. 13, 22-9.

Baumli, S., Lolli, G., Lowe, E. D., Troiani, S., Rusconi, L., Bullock, A. N., Debreczeni, J. E., Knapp, S., Johnson, L. N., 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. EMBO J. 27, 1907-18.

Bauren, G., Belikov, S., Wieslander, L., 1998. Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. Genes Dev. 12, 2759-69.

Bell, G. I., Valenzuela, P., Rutter, W. J., 1976. Phosphorylation of yeast RNA polymerases. Nature. 261, 429-31.

Berro, R., Pedati, C., Kehn-Hall, K., Wu, W., Klase, Z., Even, Y., Geneviere, A. M., Ammosova, T., Nekhai, S., Kashanchi, F., 2008. CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. J Virol. 82, 7155-66.

Bird, G., Zorio, D. A., Bentley, D. L., 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Mol Cell Biol. 24, 8963-9.

Blazek, D., 2012. The cyclin K/Cdk12 complex: an emerging new player in the maintenance of genome stability. Cell Cycle. 11, 1049-50.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., Peterlin, B. M., 2011. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev. 25, 2158-72.

Blethrow, J., Zhang, C., Shokat, K. M., Weiss, E. L., 2004. Design and use of analog-sensitive protein kinases. Curr Protoc Mol Biol. Chapter 18, Unit 18 11.

Boehm, A. K., Saunders, A., Werner, J., Lis, J. T., 2003. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. Mol Cell Biol. 23, 7628-37.

Bosken, C. A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R. P., Eick, D., Geyer, M., 2014. The structure and substrate specificity of human Cdk12/Cyclin K. Nat Commun. 5, 3505.

Bowman, E. A., Bowman, C. R., Ahn, J. H., Kelly, W. G., 2013. Phosphorylation of RNA polymerase II is independent of P-TEFb in the C. elegans germline. Development. 140, 3703-13.

Buratowski, S., 2005. Connections between mRNA 3' end processing and transcription termination. Curr Opin Cell Biol. 17, 257-61.

Buratowski, S., 2009. Progression through the RNA polymerase II CTD cycle. Mol Cell. 36, 541-6.

Burugula, B. B., Jeronimo, C., Pathak, R., Jones, J. W., Robert, F., Govind, C. K., 2014. HDACs and Phosphorylated Pol II CTD recruit Spt6 for cotranscriptional histone reassembly. Mol Cell Biol.

Cadena, D. L., Dahmus, M. E., 1987. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. J Biol Chem. 262, 12468-74.

Carey, M. F., Peterson, C. L., Smale, S. T., 2009. Dignam and Roeder nuclear extract preparation. Cold Spring Harb Protoc. 2009, pdb prot5330.

Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A., Greenleaf, A. L., 2000. Protein-interaction modules that organize nuclear function: FF domains of

CA150 bind the phosphoCTD of RNA polymerase II. Proc Natl Acad Sci U S A. 97, 9015-20.

Chao, S. H., Price, D. H., 2001. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J Biol Chem. 276, 31793-9.

Chapman, R. D., Heidemann, M., Albert, T. K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., Eick, D., 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science. 318, 1780-2.

Chapman, R. D., Heidemann, M., Hintermair, C., Eick, D., 2008. Molecular evolution of the RNA polymerase II CTD. Trends Genet. 24, 289-96.

Chen, H. H., Wang, Y. C., Fann, M. J., 2006. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. Mol Cell Biol. 26, 2736-45.

Chen, H. H., Wong, Y. H., Geneviere, A. M., Fann, M. J., 2007. CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. Biochem Biophys Res Commun. 354, 735-40.

Cheng, S. W., Kuzyk, M. A., Moradian, A., Ichu, T. A., Chang, V. C., Tien, J. F., Vollett, S. E., Griffith, M., Marra, M. A., Morin, G. B., 2012. Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. Mol Cell Biol. 32, 4691-704.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., Buratowski, S., 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev. 15, 3319-29.

Cho, E. J., Takagi, T., Moore, C. R., Buratowski, S., 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 11, 3319-26.

Cooke, C., Alwine, J. C., 1996. The cap and the 3' splice site similarly affect polyadenylation efficiency. Mol Cell Biol. 16, 2579-84.

Corden, J. L., 1990. Tails of RNA polymerase II. Trends Biochem Sci. 15, 383-7.

Corden, J. L., 2013. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. Chem Rev. 113, 8423-55.

Corden, J. L., Cadena, D. L., Ahearn, J. M., Jr., Dahmus, M. E., 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A. 82, 7934-8.

Corden, J. L., Patturajan, M., 1997. A CTD function linking transcription to splicing. Trends Biochem Sci. 22, 413-6.

Coudreuse, D., van Bakel, H., Dewez, M., Soutourina, J., Parnell, T., Vandenhaute, J., Cairns, B., Werner, M., Hermand, D., 2010. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. Curr Biol. 20, 1053-64.

Czudnochowski, N., Bosken, C. A., Geyer, M., 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat Commun. 3, 842.

Dahmus, M. E., 1981. Phosphorylation of eukaryotic DNA-dependent RNA polymerase. Identification of calf thymus RNA polymerase subunits phosphorylated by two purified protein kinases, correlation with in vivo sites of phosphorylation in HeLa cell RNA polymerase II. J Biol Chem. 256, 3332-9.

Dahmus, M. E., 1994. The role of multisite phosphorylation in the regulation of RNA polymerase II activity. Prog Nucleic Acid Res Mol Biol. 48, 143-79.

Dai, Q., Lei, T., Zhao, C., Zhong, J., Tang, Y. Z., Chen, B., Yang, J., Li, C., Wang, S., Song, X., Li, L., Li, Q., 2012. Cyclin K-containing kinase complexes maintain self-renewal in murine embryonic stem cells. J Biol Chem. 287, 25344-52.

David, C. J., Boyne, A. R., Millhouse, S. R., Manley, J. L., 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev. 25, 972-83.

Davidson, L., Muniz, L., West, S., 2014. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev. 28, 342-56.

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., Kornblihtt, A. R., 2003. A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell. 12, 525-32.

de la Mata, M., Kornblihtt, A. R., 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nat Struct Mol Biol. 13, 973-80.

Drouin, S., Laramee, L., Jacques, P. E., Forest, A., Bergeron, M., Robert, F., 2010. DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. PLoS Genet. 6, e1001173.

Du, L., Warren, S. L., 1997. A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. J Cell Biol. 136, 5-18.

Dye, M. J., Gromak, N., Proudfoot, N. J., 2006. Exon tethering in transcription by RNA polymerase II. Mol Cell. 21, 849-59.

Edwards, M. C., Wong, C., Elledge, S. J., 1998. Human cyclin K, a novel RNA polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity. Mol Cell Biol. 18, 4291-300.

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., Murphy, S., 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science. 318, 1777-9.

Egloff, S., O'Reilly, D., Murphy, S., 2008. Expression of human snRNA genes from beginning to end. Biochem Soc Trans. 36, 590-4.

Egloff, S., Szczepaniak, S. A., Dienstbier, M., Taylor, A., Knight, S., Murphy, S., 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. J Biol Chem. 285, 20564-9.

Eick, D., Geyer, M., 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev. 113, 8456-90.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., Blencowe, B. J., Ingles, C. J., 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA. 8, 1102-11.

Even, Y., Durieux, S., Escande, M. L., Lozano, J. C., Peaucellier, G., Weil, D., Geneviere, A. M., 2006. CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. J Cell Biochem. 99, 890-904.

Fabrega, C., Shen, V., Shuman, S., Lima, C. D., 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell. 11, 1549-61.

Fisher, R. P., 2005. Secrets of a double agent: CDK7 in cell-cycle control and transcription. J Cell Sci. 118, 5171-80.

Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W., Gilmartin, G. M., 1997. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A. 94, 11893-8.

Fu, T. J., Peng, J., Lee, G., Price, D. H., Flores, O., 1999. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. J Biol Chem. 274, 34527-30.

Garriga, J., Mayol, X., Grana, X., 1996. The CDC2-related kinase PITALRE is the catalytic subunit of active multimeric protein complexes. Biochem J. 319 (Pt 1), 293-8.

Garriga, J., Xie, H., Obradovic, Z., Grana, X., 2010. Selective control of gene expression by CDK9 in human cells. J Cell Physiol. 222, 200-8.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R. P., Bentley, D. L., 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol Cell Biol. 29, 5455-64.

Goldstrohm, A. C., Greenleaf, A. L., Garcia-Blanco, M. A., 2001. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. Gene. 277, 31-47.

Govind, C. K., Qiu, H., Ginsburg, D. S., Ruan, C., Hofmeyer, K., Hu, C., Swaminathan, V., Workman, J. L., Li, B., Hinnebusch, A. G., 2010. Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. Mol Cell. 39, 234-46.

Greenleaf, A. L., 1993. Positive patches and negative noodles: linking RNA processing to transcription? Trends Biochem Sci. 18, 117-9.

Grenetier, S., Bouchoux, C., Goguel, V., 2006. CTD kinase I is required for the integrity of the rDNA tandem array. Nucleic Acids Res. 34, 4996-5006.

Gu, B., Eick, D., Bensaude, O., 2013. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res. 41, 1591-603.

Guidi, B. W., Bjornsdottir, G., Hopkins, D. C., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Myers, L. C., 2004. Mutual targeting of mediator and the TFIIH kinase Kin28. J Biol Chem. 279, 29114-20.

Guo, Z., Stiller, J. W., 2004. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. BMC Genomics. 5, 69.

Harlow, E., Lane, D., 1999. Using antibodies : a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hart, R. P., McDevitt, M. A., Nevins, J. R., 1985. Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences. Cell. 43, 677-83.

Hengartner, C. J., Myer, V. E., Liao, S. M., Wilson, C. J., Koh, S. S., Young, R. A., 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell. 2, 43-53.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., Chapman, R. D., Andrau, J. C., Eick, D., 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J. 31, 2784-97.

Hirose, Y., Manley, J. L., 1998. RNA polymerase II is an essential mRNA polyadenylation factor. Nature. 395, 93-6.

Hirose, Y., Tacke, R., Manley, J. L., 1999. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev. 13, 1234-9.

Ho, C. K., Shuman, S., 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell. 3, 405-11.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., Tyers, M., 2002. Systematic

identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 415, 180-3.

Hong, S. W., Hong, S. M., Yoo, J. W., Lee, Y. C., Kim, S., Lis, J. T., Lee, D. K., 2009. Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome. Proc Natl Acad Sci U S A. 106, 14276-80.

Hsin, J. P., Li, W., Hoque, M., Tian, B., Manley, J. L., 2014. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. Elife. 3, e02112.

Hsin, J. P., Sheth, A., Manley, J. L., 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science. 334, 683-6.

Hu, Y., Raynard, S., Sehorn, M. G., Lu, X., Bussen, W., Zheng, L., Stark, J. M., Barnes, E. L., Chi, P., Janscak, P., Jasin, M., Vogel, H., Sung, P., Luo, G., 2007. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. Genes Dev. 21, 3073-84.

Huang da, W., Sherman, B. T., Lempicki, R. A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1-13.

Huang da, W., Sherman, B. T., Lempicki, R. A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4, 44-57.

Inoue, K., Ohno, M., Sakamoto, H., Shimura, Y., 1989. Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. Genes Dev. 3, 1472-9.

Islam, M. N., Fox, D., 3rd, Guo, R., Enomoto, T., Wang, W., 2010. RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. Mol Cell Biol. 30, 2460-72.

Jacobs, E. Y., Ogiwara, I., Weiner, A. M., 2004. Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. Mol Cell Biol. 24, 846-55.

Jones, J. C., Phatnani, H. P., Haystead, T. A., MacDonald, J. A., Alam, S. M., Greenleaf, A. L., 2004. C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. J Biol Chem. 279, 24957-64.

Joshi, A. A., Struhl, K., 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell. 20, 971-8.

Joshi, P. M., Sutor, S. L., Huntoon, C. J., Karnitz, L. M., 2014. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. J Biol Chem. 289, 9247-53.

Jove, R., Manley, J. L., 1984. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. J Biol Chem. 259, 8513-21.

Kadener, S., Cramer, P., Nogues, G., Cazalla, D., de la Mata, M., Fededa, J. P., Werbajh, S. E., Srebrow, A., Kornblihtt, A. R., 2001. Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. EMBO J. 20, 5759-68.

Kanagaraj, R., Huehn, D., MacKellar, A., Menigatti, M., Zheng, L., Urban, V., Shevelev, I., Greenleaf, A. L., Janscak, P., 2010. RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. Nucleic Acids Res. 38, 8131-40.

Kanin, E. I., Kipp, R. T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K. M., Ansari, A. Z., 2007. Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proc Natl Acad Sci U S A. 104, 5812-7.

Karagiannis, J., Balasubramanian, M. K., 2007. A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. PLoS One. 2, e433.

Kedinger, C., Gissinger, F., Chambon, P., 1974. Animal DNA-dependent RNA polymerases. Molecular structures and immunological properties of calf-thymus enzyme AI and of calf-thymus and rat-liver enzymes B. Eur J Biochem. 44, 421-36.

Kelly, W. G., Dahmus, M. E., Hart, G. W., 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem. 268, 10416-24.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12, 996-1006.

Keogh, M. C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., Chin, K., Punna, T., Thompson, N. J., Boone, C., Emili, A., Weissman, J. S., Hughes, T. R., Strahl, B. D., Grunstein, M., Greenblatt, J. F., Buratowski, S., Krogan, N. J., 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell. 123, 593-605.

Keogh, M. C., Podolny, V., Buratowski, S., 2003. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. Mol Cell Biol. 23, 7005-18.

Kim, E., Du, L., Bregman, D. B., Warren, S. L., 1997. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J Cell Biol. 136, 19-28.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J. H., Pollock, D. D., Megee, P. C., Bentley, D. L., 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol. 17, 1279-86.

Kim, J. B., Sharp, P. A., 2001. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem. 276, 12317-23.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., Buratowski, S., 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 432, 517-22.

Kim, M., Suh, H., Cho, E. J., Buratowski, S., 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J Biol Chem. 284, 26421-6.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., Strahl, B. D., 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol. 25, 3305-16.

Ko, T. K., Kelly, E., Pines, J., 2001. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. J Cell Sci. 114, 2591-603.

Komarnitsky, P., Cho, E. J., Buratowski, S., 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 14, 2452-60.

Krishnamurthy, S., Ghazy, M. A., Moore, C., Hampsey, M., 2009. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Mol Cell Biol. 29, 2925-34.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., Hampsey, M., 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell. 14, 387-94.

Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F., Shilatifard, A., 2003a. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell. 11, 721-9.

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., Greenblatt, J., 2003b. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol. 23, 4207-18.

Kuehner, J. N., Pearson, E. L., Moore, C., 2011. Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol. 12, 283-94.

Kwon, S. H., Florens, L., Swanson, S. K., Washburn, M. P., Abmayr, S. M., Workman, J. L., 2010. Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. Genes Dev. 24, 2133-45.

Kyburz, A., Sadowski, M., Dichtl, B., Keller, W., 2003. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. Nucleic Acids Res. 31, 3936-45.

Lapidot-Lifson, Y., Patinkin, D., Prody, C. A., Ehrlich, G., Seidman, S., Ben-Aziz, R., Benseler, F., Eckstein, F., Zakut, H., Soreq, H., 1992. Cloning and antisense oligodeoxynucleotide inhibition of a human homolog of cdc2 required in hematopoiesis. Proc Natl Acad Sci U S A. 89, 579-83.

Larochelle, S., Amat, R., Glover-Cutter, K., Sanso, M., Zhang, C., Allen, J. J., Shokat, K. M., Bentley, D. L., Fisher, R. P., 2012. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. Nat Struct Mol Biol. 19, 1108-15.

Lee, J. M., Greenleaf, A. L., 1989. A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II. Proc Natl Acad Sci U S A. 86, 3624-8.

Lee, J. M., Greenleaf, A. L., 1991. CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. Gene Expr. 1, 149-67.

Lenasi, T., Barboric, M., 2010. P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms. RNA Biol. 7, 145-50.

Lenstra, T. L., Tudek, A., Clauder, S., Xu, Z., Pachis, S. T., van Leenen, D., Kemmeren, P., Steinmetz, L. M., Libri, D., Holstege, F. C., 2013. The role of Ctk1 kinase in termination of small non-coding RNAs. PLoS One. 8, e80495.

Leroy, D., Birck, C., Brambilla, P., Samama, J. P., Ducommun, B., 1996. Characterisation of human cdc2 lysine 33 mutations expressed in the fission yeast Schizosaccharomyces pombe. FEBS Lett. 379, 217-21.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., Mattaj, I. W., 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev. 10, 1683-98.

Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., Bentley, D. L., 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell. 9, 1101-11.

Liu, J., Kipreos, E. T., 2000. Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and metazoa. Mol Biol Evol. 17, 1061-74.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., Hahn, S., 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Mol Cell Biol. 29, 4852-63.

Longley, M. J., Pierce, A. J., Modrich, P., 1997. DNA polymerase delta is required for human mismatch repair in vitro. J Biol Chem. 272, 10917-21.

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., Misteli, T., 2010. Regulation of alternative splicing by histone modifications. Science. 327, 996-1000.

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., Varani, G., 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol. 17, 1195-201.

MacKellar, A. L., Greenleaf, A. L., 2011. Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. J Biol Chem. 286, 36385-95.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., Church, G. M., 2013. RNA-guided human genome engineering via Cas9. Science. 339, 823-6.

Maniatis, T., Reed, R., 2002. An extensive network of coupling among gene expression machines. Nature. 416, 499-506.

Marques, F., Moreau, J. L., Peaucellier, G., Lozano, J. C., Schatt, P., Picard, A., Callebaut, I., Perret, E., Geneviere, A. M., 2000. A new subfamily of high molecular mass CDC2-related kinases with PITAI/VRE motifs. Biochem Biophys Res Commun. 279, 832-7.

Marshall, N. F., Peng, J., Xie, Z., Price, D. H., 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem. 271, 27176-83.

Marshall, N. F., Price, D. H., 1992. Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol. 12, 2078-90.

Marshall, N. F., Price, D. H., 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem. 270, 12335-8.

Max, T., Sogaard, M., Svejstrup, J. Q., 2007. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem. 282, 14113-20.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., Cramer, P., 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science. 336, 1723-5.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., Cramer, P., 2010. Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol. 17, 1272-8.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., Bentley, D. L., 1997a. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 11, 3306-18.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., Bentley, D. L., 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature. 385, 357-61.

Medlin, J. E., Uguen, P., Taylor, A., Bentley, D. L., Murphy, S., 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. EMBO J. 22, 925-34.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., Cramer, P., 2005. A structural perspective of CTD function. Genes Dev. 19, 1401-15.

Morris, D. P., Greenleaf, A. L., 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem. 275, 39935-43.

Morris, D. P., Lee, J. M., Sterner, D. E., Brickey, W. J., Greenleaf, A. L., 1997. Assaying CTD kinases in vitro and phosphorylation-modulated properties of RNA polymerase II in vivo. Methods. 12, 264-75.

Mortillaro, M. J., Blencowe, B. J., Wei, X., Nakayasu, H., Du, L., Warren, S. L., Sharp, P. A., Berezney, R., 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proc Natl Acad Sci U S A. 93, 8253-7.

Mosley, A. L., Pattenden, S. G., Carey, M., Venkatesh, S., Gilmore, J. M., Florens, L., Workman, J. L., Washburn, M. P., 2009. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol Cell. 34, 168-78.

Munoz, M. J., de la Mata, M., Kornblihtt, A. R., 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem Sci. 35, 497-504.

Nechaev, S., Adelman, K., 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochim Biophys Acta. 1809, 34-45.

Ni, Z., Xu, C., Guo, X., Hunter, G. O., Kuznetsova, O. V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W. H., Li, J., Young, P., Olsen, J. B., Wan, C., Loppnau, P., El

Bakkouri, M., Senisterra, G. A., He, H., Huang, H., Sidhu, S. S., Emili, A., Murphy, S., Mosley, A. L., Arrowsmith, C. H., Min, J., Greenblatt, J. F., 2014. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. Nat Struct Mol Biol. 21, 686-95.

Niwa, M., Berget, S. M., 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev. 5, 2086-95.

Noble, C. G., Hollingworth, D., Martin, S. R., Ennis-Adeniran, V., Smerdon, S. J., Kelly, G., Taylor, I. A., Ramos, A., 2005. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol. 12, 144-51.

Nonet, M., Sweetser, D., Young, R. A., 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell. 50, 909-15.

Patturajan, M., Conrad, N. K., Bregman, D. B., Corden, J. L., 1999. Yeast carboxyl-terminal domain kinase I positively and negatively regulates RNA polymerase II carboxyl-terminal domain phosphorylation. J Biol Chem. 274, 27823-8.

Patturajan, M., Schulte, R. J., Sefton, B. M., Berezney, R., Vincent, M., Bensaude, O., Warren, S. L., Corden, J. L., 1998a. Growth-related changes in phosphorylation of yeast RNA polymerase II. J Biol Chem. 273, 4689-94.

Patturajan, M., Wei, X., Berezney, R., Corden, J. L., 1998b. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. Mol Cell Biol. 18, 2406-15.

Pei, Y., Du, H., Singer, J., Stamour, C., Granitto, S., Shuman, S., Fisher, R. P., 2006. Cyclin-dependent kinase 9 (Cdk9) of fission yeast is activated by the CDK-activating kinase Csk1, overlaps functionally with the TFIIH-associated kinase Mcs6, and associates with the mRNA cap methyltransferase Pcm1 in vivo. Mol Cell Biol. 26, 777-88.

Pei, Y., Shuman, S., 2003. Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis. J Biol Chem. 278, 43346-56.

Peterlin, B. M., Price, D. H., 2006. Controlling the elongation phase of transcription with P-TEFb. Mol Cell. 23, 297-305.

Phatnani, H. P., Greenleaf, A. L., 2006. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 20, 2922-36.

Phatnani, H. P., Jones, J. C., Greenleaf, A. L., 2004. Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. Biochemistry. 43, 15702-19.

Price, D. H., Sluder, A. E., Greenleaf, A. L., 1987. Fractionation of transcription factors for RNA polymerase II from Drosophila Kc cell nuclear extracts. J Biol Chem. 262, 3244-55.

Proudfoot, N., 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol. 16, 272-8.

Proudfoot, N. J., Furger, A., Dye, M. J., 2002. Integrating mRNA processing with transcription. Cell. 108, 501-12.

Qiu, H., Hu, C., Hinnebusch, A. G., 2009. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Mol Cell. 33, 752-62.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 8, 2281-308.

Ranuncolo, S. M., Ghosh, S., Hanover, J. A., Hart, G. W., Lewis, B. A., 2012. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. J Biol Chem. 287, 23549-61.

Rasmussen, E. B., Lis, J. T., 1993. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A. 90, 7923-7.

Reyes-Reyes, M., Hampsey, M., 2007. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. Mol Cell Biol. 27, 926-36.

Rodrigues, F., Thuma, L., Klambt, C., 2012. The regulation of glial-specific splicing of Neurexin IV requires HOW and Cdk12 activity. Development. 139, 1765-76.

Rodriguez-Paredes, M., Ceballos-Chavez, M., Esteller, M., Garcia-Dominguez, M., Reyes, J. C., 2009. The chromatin remodeling factor CHD8 interacts with elongating RNA polymerase II and controls expression of the cyclin E2 gene. Nucleic Acids Res. 37, 2449-60.

Schmerwitz, U. K., Sass, G., Khandoga, A. G., Joore, J., Mayer, B. A., Berberich, N., Totzke, F., Krombach, F., Tiegs, G., Zahler, S., Vollmar, A. M., Furst, R., 2011. Flavopiridol protects against inflammation by attenuating leukocyte-endothelial interaction via inhibition of cyclin-dependent kinase 9. Arterioscler Thromb Vasc Biol. 31, 280-8.

Schor, I. E., Rascovan, N., Pelisch, F., Allo, M., Kornblihtt, A. R., 2009. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci U S A. 106, 4325-30.

Schroder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D. A., Kaehlcke, K., Cho, S., Pollard, K. S., Capra, J. A., Schnolzer, M., Cole, P. A., Geyer, M., Bruneau, B. G., Adelman, K., Ott, M., 2013. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell. 52, 314-24.

Schroeder, S. C., Schwer, B., Shuman, S., Bentley, D., 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. 14, 2435-40.

Schwartz, B. E., Werner, J. K., Lis, J. T., 2004. Indirect immunofluorescent labeling of Drosophila polytene chromosomes: visualizing protein interactions with chromatin in vivo. Methods Enzymol. 376, 393-404.

Schwartz, L. B., Roeder, R. G., 1975. Purification and subunit structure of deoxyribonucleic acid-dependent ribonucleic acid polymerase II from the mouse plasmacytoma, MOPC 315. J Biol Chem. 250, 3221-8.

Shah, K., Liu, Y., Deirmengian, C., Shokat, K. M., 1997. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. Proc Natl Acad Sci U S A. 94, 3565-70.

Sims, R. J., 3rd, Rojas, L. A., Beck, D., Bonasio, R., Schuller, R., Drury, W. J., 3rd, Eick, D., Reinberg, D., 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science. 332, 99-103.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J. D., Shaffer, S. A., Weng, Z., Moore, M. J., 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell. 151, 750-64.

Skaar, D. A., Greenleaf, A. L., 2002. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. Mol Cell. 10, 1429-39.

Skantar, A. M., Greenleaf, A. L., 1995. Identifying a transcription factor interaction site on RNA polymerase II. Gene Expr. 5, 49-69.

Sterner, D. E., Lee, J. M., Hardin, S. E., Greenleaf, A. L., 1995. The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. Mol Cell Biol. 15, 5716-24.

Stiller, J. W., Hall, B. D., 2002. Evolution of the RNA polymerase II C-terminal domain. Proc Natl Acad Sci U S A. 99, 6091-6.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., Cramer, P., 2010. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem. 285, 41597-603.

Szymczak-Workman, A. L., Vignali, K. M., Vignali, D. A., 2012. Design and construction of 2A peptide-linked multicistronic vectors. Cold Spring Harb Protoc. 2012, 199-204.

Tietjen, J. R., Zhang, D. W., Rodriguez-Molina, J. B., White, B. E., Akhtar, M. S., Heidemann, M., Li, X., Chapman, R. D., Shokat, K., Keles, S., Eick, D., Ansari, A. Z., 2010. Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol. 17, 1154-61.

Vagner, S., Vagner, C., Mattaj, I. W., 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev. 14, 403-13.

van den Heuvel, S., Harlow, E., 1993. Distinct roles for cyclin-dependent kinases in cell cycle control. Science. 262, 2050-4.

Viladevall, L., St Amour, C. V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J. J., Shokat, K. M., Schwer, B., Leatherwood, J. K., Fisher, R. P., 2009. TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. Mol Cell. 33, 738-51.

Weake, V. M., Swanson, S. K., Mushegian, A., Florens, L., Washburn, M. P., Abmayr, S. M., Workman, J. L., 2009. A novel histone fold domain-containing protein that replaces TAF6 in Drosophila SAGA is required for SAGA-dependent gene expression. Genes Dev. 23, 2818-23.

Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M., Greenleaf, A. L., 1993. Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. Genes Dev. 7, 2329-44.

Werner-Allen, J. W., Lee, C. J., Liu, P., Nicely, N. I., Wang, S., Greenleaf, A. L., Zhou, P., 2011. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. J Biol Chem. 286, 5717-26.

Westmoreland, T. J., Wickramasekara, S. M., Guo, A. Y., Selim, A. L., Winsor, T. S., Greenleaf, A. L., Blackwell, K. L., Olson, J. A., Jr., Marks, J. R., Bennett, C. B., 2009. Comparative genome-wide screening identifies a conserved doxorubicin repair network that is diploid specific in Saccharomyces cerevisiae. PLoS One. 4, e5830.

Wilcox, C. B., Rossettini, A., Hanes, S. D., 2004. Genetic interactions with C-terminal domain (CTD) kinases and the CTD of RNA Pol II suggest a role for ESS1 in transcription initiation and elongation in Saccharomyces cerevisiae. Genetics. 167, 93-105.

Winsor, T. S., Bartkowiak, B., Bennett, C. B., Greenleaf, A. L., 2013. A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II. PLoS One. 8, e60909.

Wood, A., Shilatifard, A., 2006. Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. Cell Cycle. 5, 1066-8.

Xiang, K., Manley, J. L., Tong, L., 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. Nat Commun. 3, 946.

Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M. M., Manley, J. L., Tong, L., 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature. 467, 729-33.

Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., Strahl, B. D., 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. 17, 654-63.

Xu, Y. X., Manley, J. L., 2007. Pin1 modulates RNA polymerase II activity during the transcription cycle. Genes Dev. 21, 2950-62.

Yao, S., Neiman, A., Prelich, G., 2000. BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo. Mol Cell Biol. 20, 7080-7.

Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., Jones, K. A., 2007. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev. 21, 160-74.

Yoh, S. M., Lucas, J. S., Jones, K. A., 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev. 22, 3422-34.

Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., Mellor, J., 2008. Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol Cell Biol. 28, 4915-26.

Yu, D. S., Cortez, D., 2011. A role for CDK9-cyclin K in maintaining genome integrity. Cell Cycle. 10, 28-32.

Yu, D. S., Zhao, R., Hsu, E. L., Cayer, J., Ye, F., Guo, Y., Shyr, Y., Cortez, D., 2010. Cyclin-dependent kinase 9-cyclin K functions in the replication stress response. EMBO Rep. 11, 876-82.

Yu, J., Pacifico, S., Liu, G., Finley, R. L., Jr., 2008. DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics. 9, 461.

Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., Shatkin, A. J., 1997. Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. Proc Natl Acad Sci U S A. 94, 12898-903.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., Corden, J. L., 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc Natl Acad Sci U S A. 93, 6975-80.

Zehring, W. A., Lee, J. M., Weeks, J. R., Jokerst, R. S., Greenleaf, A. L., 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. Proc Natl Acad Sci U S A. 85, 3698-702.

Zeng, C., Berget, S. M., 2000. Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol Cell Biol. 20, 8290-301.

Zhang, D. W., Mosley, A. L., Ramisetty, S. R., Rodriguez-Molina, J. B., Washburn, M. P., Ansari, A. Z., 2012. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J Biol Chem. 287, 8541-51.

Zhang, J., Corden, J. L., 1991. Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. J Biol Chem. 266, 2290-6.

Zhong, X. Y., Wang, P., Han, J., Rosenfeld, M. G., Fu, X. D., 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Mol Cell. 35, 1-10.

Biography

Bartlomiej (Bart) Bartkowiak was born in Poznań, Poland on August 26, 1985 to parents Magdalena and Miroslaw Bartkowiak. From 2003-2007 he attended the University of California Santa Barbara where he earned a B.S. in Biochemistry and Molecular Biology. After graduation he pursued his passion for the biological sciences by becoming a research technician in the laboratory of Dr. Rasheed Gbadegesin in the Center for Human Genetics at Duke University. He entered the doctoral program in Biochemistry at Duke University in the fall of 2008 and has been working towards his degree in the laboratory of Dr. Arno Greenleaf. His publications include the following:

Bartkowiak, B., Greenleaf, A. L., 2014. Expression, Purification, and Identification of Associated Proteins of the Full Length hCDK12/CyclinK Complex. J Biol Chem. Submitted.

Winsor, T. S., **Bartkowiak, B.**, Bennett, C. B., Greenleaf, A. L., 2013. A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II. PLoS One. 8, e60909.

Gbadegesin, R. A., Brophy, P. D., Adeyemo, A., Hall, G., Gupta, I. R., Hains, D., **Bartkowiak, B.**, Rabinovich, C. E., Chandrasekharappa, S., Homstad, A., Westreich, K., Wu, G., Liu, Y., Holanda, D., Clarke, J., Lavin, P., Selim, A., Miller, S., Wiener, J. S., Ross, S. S., Foreman, J., Rotimi, C., Winn, M. P., 2013. TNXB mutations can cause vesicoureteral reflux. J Am Soc Nephrol. 24, 1313-22.

Gbadegesin, R. A., Lavin, P. J., Hall, G., **Bartkowiak, B.**, Homstad, A., Jiang, R., Wu, G., Byrd, A., Lynn, K., Wolfish, N., Ottati, C., Stevens, P., Howell, D., Conlon, P., Winn, M. P., 2012. Inverted formin 2 mutations with variable expression in patients with sporadic and hereditary focal and segmental glomerulosclerosis. Kidney Int. 81, 94-9.

Bartkowiak, B., Greenleaf, A. L., 2011. Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? Transcription. 2, 115-119.

Bartkowiak, B.*, Mackellar, A. L.*, Greenleaf, A. L., 2011. Updating the CTD Story: From Tail to Epic. Genet Res Int. 2011, 623718.

Bartkowiak, B.*, Liu, P*., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., Greenleaf, A. L., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24, 2303-16.

Gbadegesin, R., Lavin, P., Janssens, L., **Bartkowiak, B.**, Homstad, A., Wu, G., Bowling, B., Eckel, J., Potocky, C., Abbott, D., Conlon, P., Scott, W. K., Howell, D., Hauser, E., Winn, M. P., 2010. A new locus for familial FSGS on chromosome 2p. J Am Soc Nephrol. 21, 1390-7.

Gbadegesin, R., **Bartkowiak, B.**, Lavin, P. J., Mukerji, N., Wu, G., Bowling, B., Eckel, J., Damodaran, T., Winn, M. P., 2009. Exclusion of homozygous PLCE1 (NPHS3) mutations in 69 families with idiopathic and hereditary FSGS. Pediatr Nephrol. 24, 281-5.

* These authors contributed equally